Contents

Chapter 2

Elementary probability theory

Fundamental to the understanding of the Monte Carlo method and interpretion of its results, is a basic understanding of elementary probability theory. In this chapter we introduce some elementary probability theory to facilitate later discussions, and establish some notation.

2.1 Continuous random variables

2.1.1 One-dimensional probability distributions

A probability distribution function on x, $p(x)$, also known as a "pdf", or "PDF", is a measure of the likelihood of observing x over some range, $x_{\rm min} \le x \le x_{\rm max}$. For example, if x is the distance from its point of creation, at which a photon interacts via the Compton interaction, the statement $p(x_1) = 2p(x_2)$ means that an observation of x in a differential interval $x_1 \le x \le x_1 + {\rm d}x$ is twice as likely to be observed than in an interval $x_2 \le x \le x_2 + {\rm d}x$, in the limit that ${\rm d}x$ goes to zero. An example pdf, $p(x) = \exp(-x)$, is shown in Figure 2.1.

A pdf has necessary properties:

- $p(x) \ge 0 \ \forall \ x_{\rm min} \le x \le x_{\rm max}$.
  Negative probabilities have no interpretation in our context.

- $p(x)$ must be normalizable, and is normalized in the following fashion:

$$\int_{x_{\rm min}}^{x_{\rm max}} {\rm d}x \ p(x) = 1 \ . \tag{2.1}$$

- $-\infty < x_{\rm min} < x_{\rm max} < +\infty$, that is, $x_{\rm min}$ and $x_{\rm max}$ can be any real number, including $\pm\infty$, so long as $x_{\rm min} < x_{\rm max}$.

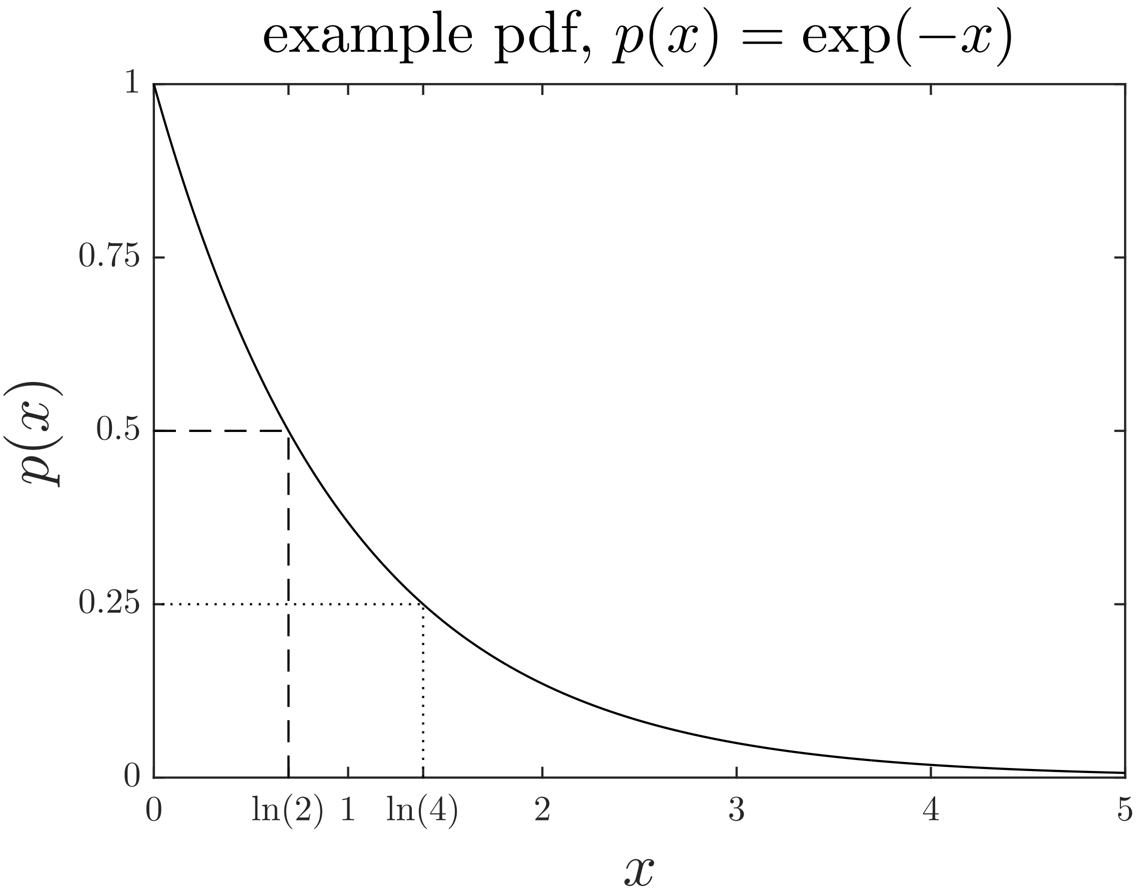

Figure 2.1: The pdf $p(x) = \exp(-x)$. $p(\ln 2) = 2p(\ln 4)$

These are the only restrictions on $p(x)$. Note that the statement of normalization above implies that $p(x)$ is integrable over its range of definition. The pdf may be discontinuous and even infinite. For example, $p(x) = a\delta(x - x_1) + (1 - a)\delta(x - x_2)\ \forall\ |x| < \infty$, where $\delta()$ is the Dirac delta function, is an example of a properly defined pdf.

Cumulative distribution functions

Associated with every pdf, is its cumulative distribution function, $c(x)$, also known as the "cdf", or "CDF" The cdf is computed below:

$$c(x) = \int_{x_{\min}}^{x} \mathrm{d}x\ p(x)\ , \tag{2.2}$$

and has the property:

$$c'(x) = p(x)\ . \tag{2.3}$$

The cdf has a critical role to play in Monte Carlo methods. In this chapter we shall illustrate by example.

Moments of probability distribution functions

Certain probability functions can be characterized in terms of their integer moments,

$$\begin{aligned} \langle x^n \rangle &= \int_{x_{\min}}^{x_{\max}} \mathrm{d}x\ x^n p(x)\ \forall\ n \geq 0 \wedge n \in \mathbb{N}\ , \\ \langle x^0 \rangle &= 1\ (\text{by definition})\ . \end{aligned} \tag{2.4}$$

However, the existence of these moments is not guaranteed nor even necessary. When $\langle x \rangle$ does exist, it is given the symbol μ, the average value of the probability distribution. When both $\langle x \rangle$ and $\langle x^2 \rangle$ exist, the variance associated with the probability function may be defined to be:

$$\mathrm{var}\{x\} = \sigma^2 = \langle x^2 \rangle - \langle x \rangle^2\ , \tag{2.5}$$

where σ is a measure of the "width" of the probability distribution, measured in the same units as μ. $\mathrm{var}\{x\}$ is zero for the Dirac delta function and greater than zero for all other probability distribution functions, even combinations of delta functions.

Examples of probability distributions

The δ-function probability distribution

Consider the probability distribution,

$$p(x) = \delta(x - x_0)\ \forall\ |x| < \infty\ . \tag{2.6}$$

By definition of the delta function,

$$\begin{aligned} \int_{-\infty}^{\infty} \mathrm{d}x\, p(x) &= 1\ , \\ \int_{-\infty}^{\infty} \mathrm{d}x\, x^n\, p(x) &= x_0^n\ \forall\ n \geq 0 \\ c(x) = \Theta(x - x_0)\ . \end{aligned}$$

where $\Theta()$ is the Heaviside step function.
Note that $\mu = x_0$ and $\sigma = \sqrt{\langle x^2 \rangle - \langle x \rangle^2} = 0$ in this case. The pdf and cdf are shown in 2.2.

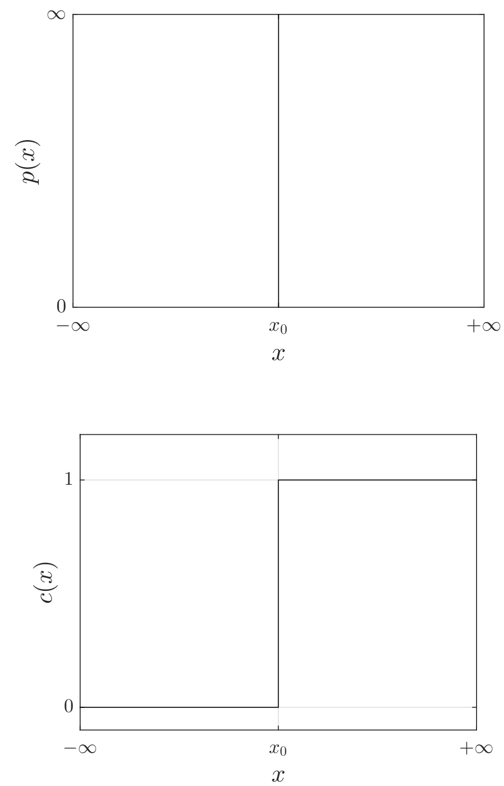

Figure 2.2: The pdf $p(x) = \delta(x - x_0)$ and its associated cdf.

The "true or false" probability distribution

Consider the "true or false" probability distribution,

$$p(x) = (1/2)[\delta(x) + \delta(x-1)] \;\forall\; |x| < \infty. \tag{2.7}$$

This could describe a coin-flip experiment where a result of zero (or false) corresponds to "tails" and a result of one (or true) correspondes to "heads".

By definition of the delta function,

$$\begin{aligned}
\int_{-\infty}^{\infty} \mathrm{d}x\, p(x) &= 1\ , \\
\int_{-\infty}^{\infty} \mathrm{d}x\, x^n\, p(x) &= \frac{1}{2} \;\forall\; n > 0\ , \\
c(x) &= \frac{1}{2}[\Theta(x) + \Theta(x-1)]\ .
\end{aligned}$$

Note that $\mu = \frac{1}{2}$ and $\sigma = \sqrt{\langle x^2 \rangle - \langle x \rangle^2} = \frac{1}{2}$ in this case. The pdf and cdf are shown in 2.3.

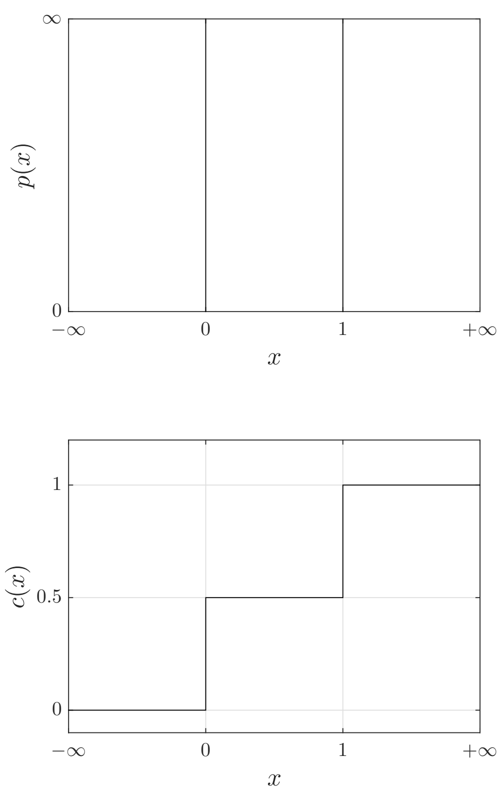

Figure 2.3: The pdf $p(x) = (1/2)[\delta(x) + \delta(x-1)]$ and its associated cdf.

The exponential probability distribution

Consider the exponential probability distribution,

$$p(x) = \Sigma \exp(-\Sigma x) \text{ for } 0 \leq x < \infty \ . \tag{2.8}$$

This probability distribution is used in the decay of nuclides and determination of pathlength distributions in particle interactions in matter. For interaction pathlength distributions, Σ is the macroscopic cross section in an attenuating material.

Computing the characteristics of this distribution,

$$\begin{aligned}
\int_0^\infty \mathrm{d}x\, p(x) &= 1 \ , \\
\int_0^\infty \mathrm{d}x\, x^n\, p(x) &= \frac{n!}{\Sigma^n}\ \forall\ n \geq 0 \ , \\
c(x) &= 1 - \exp(-\Sigma x) \ .
\end{aligned}$$

Note that $\mu = \sigma = \Sigma^{-1}$ in this case. The pdf and cdf are shown in 2.4.

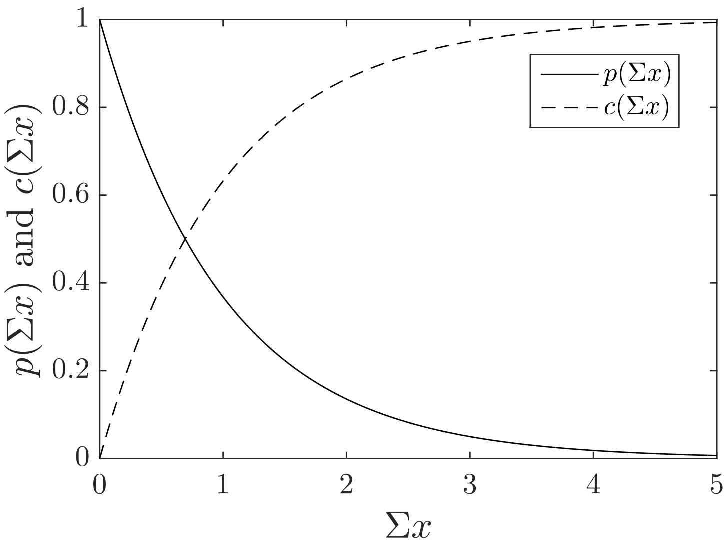

Figure 2.4: The pdf $p(\Sigma x) = \exp(-\Sigma x)$ and its associated cdf.

The Cauchy probability distribution

An example of a probability distribution function that has no moments is the Cauchy probability distribution, also known as the Lorentz, or Breit-Wigner distribution. In our context, this distribution arises from an interesting application—the intrinsic probability distribution of the energy of a quantum from an excited atomic state of finite lifetime.

The Cauchy probability distribution is

$$p(x) = \frac{1}{\pi} \frac{\gamma}{\gamma^2 + (x - x_0)^2} \ \forall \ |x| < \infty \ . \tag{2.9}$$

The Cauchy distribution has no moments, but its mode and median are x_0. The parameter γ is the half width at half maximum, since $p(x_0 \pm \gamma) = \frac{1}{2} p(x_0)$, and serves as a measure of the width of the pdf.

The mean value of the Cauchy/Lorentz distribution can be made to be x_0 if the "principle value" of the first moment is obtained in the following way:

$$\begin{aligned}
\langle x \rangle &= \frac{\gamma}{\pi} \lim_{a \longrightarrow \infty} \int_{-a-x_0}^{a-x_0} \mathrm{d}x \, \frac{x}{\gamma^2 + (x - x_0)^2} \\
&= \frac{\gamma}{\pi} \lim_{a \longrightarrow \infty} \int_{-a-x_0}^{a-x_0} \mathrm{d}x \, \frac{x_0 + (x - x_0)}{\gamma^2 + (x - x_0)^2} \\
&= x_0 + \overbrace{\frac{\gamma}{\pi} \lim_{a \longrightarrow \infty} \int_{-a}^{a} \mathrm{d}u \, \frac{u}{1 + u^2}}^{0} \text{ (by antisymmetry)}
\end{aligned} \tag{2.10}$$

but the second moment and hence the variance can not be defined in any fashion.

Of course, this procedure would cause a mathematician to have conniptions! However, infinities in physical applications can usually be rationalized by physical argument.

The Cauchy pdf, being a proper probaility function, has the cdf:

$$c(x) = \frac{1}{2} + \frac{\arctan(\frac{x-x_0}{\gamma})}{\pi} \,. \tag{2.11}$$

The Cauchy pdf and cdf for $x_0 = 0$ and $\gamma = 1$ are shown in 2.5. The pdf was denormalized for display purposes.

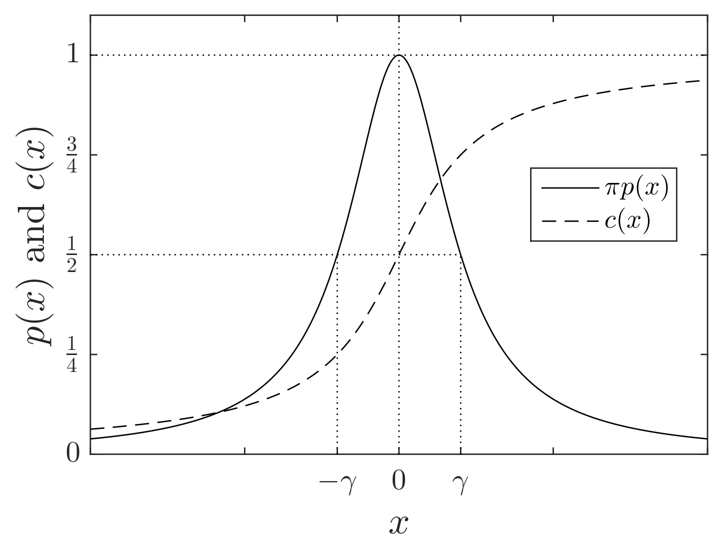

Figure 2.5: The Cauchy pdf and its associated cdf for $x_0 = 0$ and $\gamma = 1$.

<u>A generalized Cauchy probability distribution</u>

A generalized Cauchy probability distribution[1] is

$$\begin{aligned} p_m(x) &= \frac{\Gamma(m)}{\gamma\sqrt{\pi}\,\Gamma(m-1/2)}\left[\frac{\gamma^2}{\gamma^2+(x-x_0)^2}\right]^m \forall\ |x| < \infty \wedge m > \frac{1}{2} \\ c_m(x) &= \frac{1}{2} + \frac{\Gamma(m)}{\sqrt{\pi}\,\Gamma(m-1/2)}\left(\frac{x-x_0}{\gamma}\right)\,{}_2F_1\left[\frac{1}{2}, m; \frac{3}{2}; -\left(\frac{x-x_0}{\gamma}\right)^2\right] \,, \end{aligned} \tag{2.12}$$

[1]A standard version of the generalized Cauchy distribution does not yet exist, though there is some literature to be found on this topic. For this book, we will consider this form, and exploit it to illustrate aspects of moments.

where the properties of the gamma function that would be used in the evaluation of (2.12) for $m \geq 1$ and half-integral are:

$$\begin{aligned}
\Gamma(m+1) &= n\Gamma(n) \quad \text{recursive formula for } \Gamma() \\
\Gamma(1/2) &= \sqrt{\pi} \\
\Gamma(1) &= 1 \\
\Gamma(m) &= (m-1)! \\
\Gamma(m+1) &= m! \\
\Gamma(2m-1) &= \sqrt{\pi}\frac{1\cdot 3\cdot 3\cdots(2m-1)}{2^m} \\
(2m-1)!! &= \sqrt{\pi}\frac{(2m-1)!!}{2^m}
\end{aligned}$$

The generalized Cauchy distribution has its mode and median are x_0. There exist exactly $\lfloor 2(m-1)\rfloor$ moments for this distribution.[2] After centering the distribution, $u = x - x_0$,

$$\begin{aligned}
p_m(u) &= \frac{\Gamma(m)}{\gamma\sqrt{\pi}\,\Gamma(m-1/2)}\left[\frac{1}{\gamma^2+u^2}\right]^m \forall\ |u| < \infty \wedge m > \frac{1}{2} \\
c_m(u) &= \frac{1}{2} + \frac{\Gamma(m)}{\sqrt{\pi}\,\Gamma(m-1/2)}\left(\frac{u}{\gamma}\right)\,{}_2F_1\left[\frac{1}{2}, m; \frac{3}{2}; -\left(\frac{u}{\gamma}\right)^2\right] \\
\langle u^n\rangle &= \frac{\gamma^n[1+(-1)^n]\Gamma\left(\frac{n+1}{2}\right)\Gamma\left(m-\frac{n+1}{2}\right)}{2\sqrt{\pi}\,\Gamma\left(m-\frac{1}{2}\right)} \forall\ n < 2m-1 \qquad (2.13)
\end{aligned}$$

$$(2.14)$$

From (2.14) one can readily compute the moments of x using:

$$\langle u^n\rangle = \langle (x-x_0)^n\rangle \qquad (2.15)$$

[2] $\lfloor 2(m-1)\rfloor$ rounds to an integer in the direction of $-\infty$.

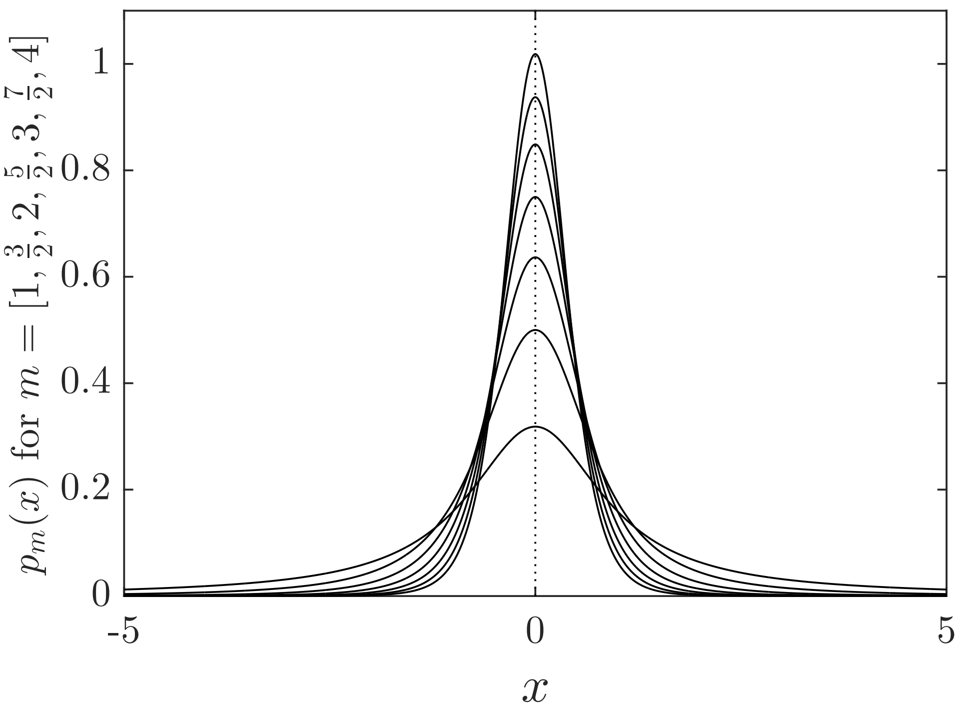

Figure 2.6: The generalized Cauchy pdf for $x_0 = 0$, $\gamma = 1$ and $m = [1, \frac{3}{2}, 2, \frac{5}{2}, 3, \frac{7}{2}, 4]$. As m increases, the distribution narrows.

The Rutherfordian probability distribution

Another interesting probability distribution function of great importance in the simulation of electron transport is the screened Rutherford or Wentzel distribution function:

$$p(\mu) = \frac{a(2+a)}{2}\frac{1}{(1-\mu+a)^2} \quad ; \quad -1 \le \mu \le 1, \tag{2.16}$$

where μ is the cosine of the scattering angle, $\cos\Theta$.

The Rutherfordian probability distribution (small-angle form)

The conventional small-angle form of the screened Rutherford or Wentzel distribution function is:

$$p(\Theta) = 4a\frac{\Theta}{(\Theta^2+2a)^2} \quad ; \quad 0 \le \Theta < \infty, \tag{2.17}$$

Its first moment exists, $\langle\Theta\rangle = \pi\sqrt{a/2}$ but its second moment is infinite! This strange behavior, the nonexistence of an angular variance is responsible for the theory of electron transport being so problematic. Of course, one could restrict the range of integration to physical angles, $0 \le \Theta \le \pi$, but the problems persist.

2.1.2 Two-dimensional probability distributions

Consideration of two and higher-dimensional probability distributions follows from a generalization of one-dimensional distributions with the added features of correlation between observables and conditional probabilities.

Consider a two-dimensional (or joint) probability function p(x,y). A tangible example is the distribution in energy and angle of a photon undergoing an inelastic collision with an atom[3].

Another example is the two-dimensional probability distribution presented in Figure 2.7. The meaning of two-dimensional probability distributions is as follows: Hold one variable, say x fixed, and the resulting distribution is a probability distribution function in the other variable, y. It as if you cut through the two-dimensional probability distribution at a given point in x and then displayed the “cross cuts”. Several examples of these cross cuts are shown in Figure 2.8.

The notions of normalization and moments follow directly. Thus,

$$\langle x^n y^m\rangle = \int_{y_{\min}}^{y_{\max}}\int_{x_{\min}}^{x_{\max}} \mathrm{d}x\mathrm{d}y\ x^n y^m p(x,y)\ , \tag{2.18}$$

[3] “Inelastic” in this definition relates to the energy of the photon changing, not the energy of the atom.

Figure 2.7: A two-dimensional probability function $f(x, y)$

with the normalization condition $\langle x^0 y^0 \rangle = 1$, the only "moment" that must be defined. Higher order moments may or may not exist. If they exist, we define the covariance:

$$\text{cov}\{x, y\} = \langle xy \rangle - \langle x \rangle \langle y \rangle \ , \tag{2.19}$$

which can be positive or negative. Note that $\text{cov}\{x, x\} = \text{var}\{x\}$.

The covariance is a measure of the independence of observing x or y. If x and y are independent random variables, then $p(x, y) = p_1(x)p_2(y)$ and $\text{cov}\{x, y\} = 0$. A related function is the correlation coefficient:

$$\rho\{x, y\} = \frac{\text{cov}\{x, y\}}{\sqrt{\text{var}\{x\}\text{var}\{y\}}} \ , \tag{2.20}$$

where $-1 \leq \rho\{x, y\} \leq 1$.

Some interesting relations involving the variances and the covariances may be found. For example,

$$\text{var}\{x \pm y\} = \text{var}\{x\} + \text{var}\{y\} \pm 2 \ \text{cov}\{x, y\} \ , \tag{2.21}$$

Figure 2.8: Cross cuts of the two-dimensional probability function given in Figure 2.7.

or simply $\mathrm{var}\{x\} + \mathrm{var}\{y\}$ if x and y are independent.

The marginal probabilities are defined by integrating out the other variables.

$$m(x) = \int_{y_{\min}}^{y_{\max}} \mathrm{d}y \; p(x,y) \quad ; \quad m(y) = \int_{x_{\min}}^{x_{\max}} \mathrm{d}x \; p(x,y) \; . \tag{2.22}$$

Note that the marginal probability distributions are properly normalized. For the example of joint energy and angular distributions, one marginal probability distribution relates to the distribution in energy irrespective of angle and the other refers to the distribution in angle irrespective of energy. This the joint probability distribution function may be written:

$$p(x,y) = m(x)p(y|x) \; , \tag{2.23}$$

where the *conditional* probability is defined by

$$p(y|x) = \frac{p(x,y)}{m(x)} \; . \tag{2.24}$$

The interpretation of the *conditional* probability is that given x, what is the probability that y occurs. The appearance of $m(x)$ in the denominator guaranteed the normalization of $p(y|x)$.

2.1.3 Cumulative probability distributions

Associated with each one-dimensional probability distribution function is its cumulative probability distribution function

$$c(x) = \int_{x_{\rm min}}^{x} {\rm d}x' \; p(x') \; . \tag{2.25}$$

Cumulative probability distribution functions have the following properties which follow directly from its definition and the properties of probability distribution functions:

- $p(x)$ and $c(x)$ are related by a derivative:

$$p(x) = \frac{{\rm d}c(x)}{{\rm d}x} \; , \tag{2.26}$$

- $c(x)$ is zero at the beginning of its range of definition

$$c(x_{\rm min}) = 0 \; , \tag{2.27}$$

- and unity at the end of its range of definition

$$c(x_{\rm max}) = 1 \; , \tag{2.28}$$

- $c(x)$ is a monotonically increasing function of x as a result of $p(x)$ always being positive and the definition of $c(x)$ in Equation 2.25.

Cumulative probability distribution functions can be related to uniform random numbers to provide a way for sampling these distributions. We will complete this discussion in Chapter **??**.

Cumulative probability distribution functions for multi-dimensional probability distribution functions are usually defined in terms of the one-dimensional forms of the marginal and conditional probability distribution functions.

2.2 Discrete random variables

A more complete discussion of probability theory would include some discussion of discrete random variables. An example would be the results of flipping a coin or a card game. We will have some small use for this in Chapter **??** and will introduce what we need at that point.

Bibliography

[BHNR94] A. F. Bielajew, H. Hirayama, W. R. Nelson, and D. W. O. Rogers. History, overview and recent improvements of EGS4. *National Research Council of Canada Report PIRS-0436*, 1994.

[Bie94] A. F. Bielajew. Monte Carlo Modeling in External Electron-Beam Radiotherapy — Why Leave it to Chance? *In "Proceedings of the XI'th Conference on the Use of Computers in Radiotherapy" (Medical Physics Publishing, Madison, Wisconsin)*, pages 2 – 5, 1994.

[Bie95] A. F. Bielajew. EGS4 timing benchmark results: Why Monte Carlo is a viable option for radiotherapy treatment planning. *In "Proceedings of the International Conference on Mathematics and Computations, Reactor Physics, and Environmental Analyses" (American Nuclear Society Press, La Grange Park, Illinois, U.S.A.)*, pages 831 – 837, 1995.

[BR92] A. F. Bielajew and D. W. O. Rogers. A standard timing benchmark for EGS4 Monte Carlo calculations. *Medical Physics*, 19:303 – 304, 1992.

[BW91] A. F. Bielajew and P. E. Weibe. EGS-Windows - A Graphical Interface to EGS. *NRCC Report: PIRS-0274*, 1991.

[dB77] G. Comte de Buffon. *Essai d'arithmétique morale*, volume 4. Supplément à l'Histoire Naturelle, 1777.

[KW86] M. H. Kalos and P. A. Whitlock. *Monte Carlo methods, Volume I: Basics*. John Wiley and Sons, New York, 1986.

[Lap86] P. S. Laplace. Theorie analytique des probabilités, Livre 2. In *Oeuvres complétes de Laplace*, volume 7, Part 2, pages 365 – 366. L'académie des Sciences, Paris, 1886.

[MRR95] M. S. MacPherson, C. K. Ross, and D. W. O. Rogers. A technique for accurate measurement of electron stopping powers. *Med. Phys. (abs)*, 22:950, 1995.

[MRR96] M. S. MacPherson, C. K. Ross, and D. W. O. Rogers. Measured electron stopping powers for elemental absorbers. *Med. Phys. (abs)*, 23:797, 1996.

[NBRH94] W. R. Nelson, A. F. Bielajew, D. W. O. Rogers, and H. Hirayama. EGS4 in '94: A decade of enhancements. *Stanford Linear Accelerator Report SLAC-PUB-6625 (Stanford, Calif)*, 1994.

[NHR85] W. R. Nelson, H. Hirayama, and D. W. O. Rogers. The EGS4 Code System. Report SLAC–265, Stanford Linear Accelerator Center, Stanford, Calif, 1985.

Problems

1. Which of the following are candidate probability distributions? For those that are not, explain. For those that are, determine the normalization constant N. Those that are proper probability distributions, accompanied by mathematical proof, which contain moments that do not exist?

   (a) $f(x) = N \exp(-\mu x) \quad 0 \le x < \infty$
   (b) $f(x) = N \exp(-\mu x) \quad 0 \le x < \Lambda/\mu$ where μ, Λ are positive, real constants
   (c) $f(x) = N \sin(x) \quad 0 \le x < \pi$
   (d) $f(x) = N \sin(x) \quad 0 \le x < 2\pi$
   (e) $f(x) = N/\sqrt{x} \quad 0 \le x < 1$
   (f) $f(x) = N/\sqrt{x} \quad 1 \le x < \infty$
   (g) $f(x) = Nx/(x^2 + a^2)^{3/2} \quad 0 \le x < \infty$

2. Verify that Equations 2.9, 2.16, and 2.17 are true probability distributions.

3. Consider the probability distribution,

$$p(x) = (1/2)[\delta(x - a) + \delta(x - b)] \ \forall \ |x| < \infty.$$

   What are all the moments of this distribution?

4. Consider the probability distribution,

$$p(x) = N[\Theta(x - a) - \Theta(x - b)] \ . \ \forall \ |x| < \infty.$$

   Can this be a proper pdf? If so, what is N. Does it matter what the relative values of a and b are? What are all the moments of this distribution?

5. Prove Equation 2.21. Simplify in the case that x and y are independent.