ECE 302: Probability and Applications[1]

Week 4 Topics

- Discrete Random Variables
  - Probability Mass Function
- Continuous Random Variables
  - Probability Density Function
- Expected Value: Mean and Variance of Random Variables

1 Discrete Random Variables and Probability Mass Function

A **discrete random variable $\boldsymbol{X}$** assumes values from a countable set, that is, $S_X = \{x_1, x_2, x_3, \ldots\}$. A discrete random variable is **finite** if its range is finite, that is, $S_X = \{x_1, x_2, \ldots, x_n\}$.

- To find the probabilities of events involving a discrete random variable X, we only need to obtain the probabilities for the events $A_k = \zeta : X(\zeta) = x_k$ in the underlying random experiment, as shown in Fig. 1.
- The probabilities of events involving X can be found from the probabilities of the A_k's.

The **probability mass function (pmf) of a discrete random variable $\boldsymbol{X}$** is defined as:

$$p_X(x) = P[X = x] = \mathrm{P}[\{\zeta : X(\zeta) = x\}] \text{ for } x \text{ a real number.} \tag{1}$$

- Note that $p_X(x)$ is a function of x over the real line, and that $p_X(x)$ is nonzero only at the values $x_1, x_2, x_3, \ldots$.
- For x_k in S_X, we have $p_X(x_k) = P[A_k]$.
- The events $A_1, A_2, \ldots$ form a partition of S as shown in Fig. 1.

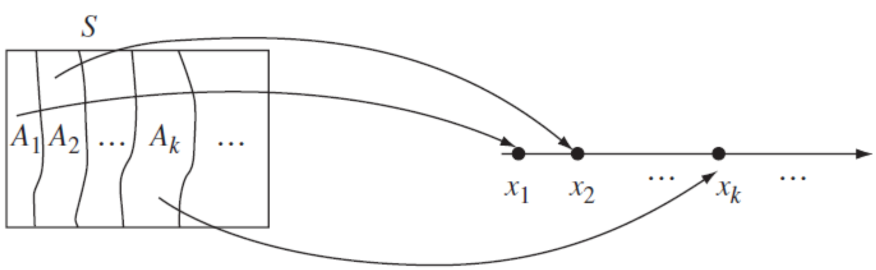


Figure 1: Partition of sample space S associated with a discrete random variable.

- To see this, we first note that for $j \neq k$, then

$$A_j \cap A_k = \{\zeta : X(\zeta) = x_j \text{ and } X(\zeta) = x_k\} = \varnothing$$

since each ζ is mapped into one and only one value in S_X.

- Next we note that S is the union of the A_k's since every ζ in S is mapped into some x_k and so belongs to some event A_k in the partition.

We can find the probability of any event involving the random variable X by expressing it as the union of events A_k's. For example, suppose for the event $B = \{x_2, x_5\}$, then

$$\begin{aligned} P[X \text{ in } B] &= P\left[\{\zeta : X(\zeta) = x_2\} \cup \{\zeta : X(\zeta) = x_5\}\right] \\ &= P\left[A_2 \cup A_5\right] = P\left[A_2\right] + P\left[A_5\right] \\ &= p_X(2) + p_X(5). \end{aligned}$$

The pmf $p_X(x)$ satisfies three properties that provide all the information required to calculate probabilities for events involving the discrete random variable X:

$$\text{(i) } p_X(x) \geq 0 \text{ for all } x \tag{2}$$

$$\text{(ii) } \sum_{x \in S_X} p_X(x) = \sum_{\text{all } k} p_X\left(x_k\right) = \sum_{\text{all } k} P\left[A_k\right] = 1 \tag{3}$$

$$\text{(iii) } P[X \text{ in } B] = \sum_{x \in B} p_X(x) \text{ where } B \subset S_X. \tag{4}$$

- If we are only interested in events concerning X, then we can forget about the underlying random experiment and its associated probability law and just work with S_X and the pmf of X.

Example 1: Coin Tosses and Binomial Random Variable

Let X be the number of heads in three independent tosses of a coin. Find the pmf of X.

Proceeding as in **Example 3.3**, we find:

$$p_0 = P[X = 0] = P[\{\text{TTT}\}] = (1 - p)^3$$

$$p_1 = P[X = 1] = P[\{\text{HTT}\}] + P[\{\text{THT}\}] + P[\{\text{TTH}\}] = 3(1 - p)^2 p$$
$$p_2 = P[X = 2] = P[\{\text{HHT}\}] + P[\{\text{HTH}\}] + P[\{\text{THH}\}] = 3(1 - p)p^2$$
$$p_3 = P[X = 3] = P[\{\text{HHH}\}] = p^3.$$

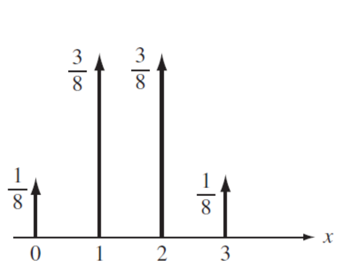
Example 2: A Betting Game

A player receives \$1 if the number of heads in three coin tosses is 2, \$8 if the number is 3 , but nothing otherwise. Find the pmf of the reward Y.

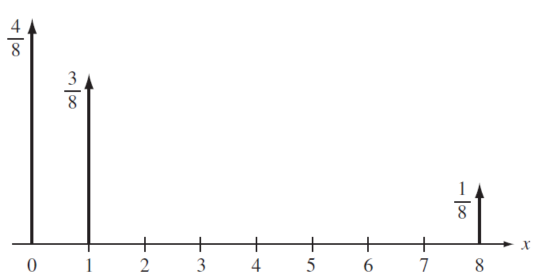
$$p_Y(0) = P[\zeta \in \{\text{TTT}, \text{TTH}, \text{THT}, \text{HTT}\}] = 4/8 = 1/2$$
$$p_Y(1) = P[\zeta \in \{\text{THH}, \text{HTH}, \text{HHT}\}] = 3/8$$
$$p_Y(8) = P[\zeta \in \{\text{HHH}\}] = 1/8$$

Note that $p_Y(0) + p_Y(1) + p_Y(8) = 1$.

Figures 2(a) and (b) show the graph of $p_X(x)$ versus x for the random variables in Examples 1 and 2, respectively. In general, the graph of the pmf of a discrete random variable has vertical arrows of height $p_X(x_k)$ at the values x_k in S_X. We may view the total probability as one unit of mass and $p_X(x)$ as the amount of probability mass that is placed at each of the discrete points $x_1, x_2, \ldots$.The relative values of pmf at different points give an indication of the relative likelihoods of occurrence.



(a) Graph of pmf in 3 coin tosses



(b) Graph of pmf in betting game

Figure 2

Example 3: Random Number Generator

A random number generator produces an integer number X that is equally likely to be any element in the set $S_X = \{0, 1, 2, \ldots, M-1\}$. Find the pmf of X.

For each k in S_X, we have $p_X(k) = 1/M$. Note that

$$p_X(0) + p_X(1) + \ldots + p_X(M-1) = 1$$

We call X the **uniform random variable** in the set $\{0, 1, \ldots, M-1\}$.

Example 4: Bernoulli Random Variable

Let A be an event of interest in some random experiment, e.g., a device is not defective. We say that a "success" occurs if A occurs when we perform the experiment. The Bernoulli random variable I_A is equal to 1 if A occurs and zero otherwise, and is given by the *indicator function for* A:

$$I_A(\zeta) = \begin{cases} 0 & \text{if } \zeta \text{ not in } A \\ 1 & \text{if } \zeta \text{ in } A \end{cases}$$

Find the pmf of I_A.

$I_A(\zeta)$ is a finite discrete random variable with values from $S_I = \{0, 1\}$, with pmf:

$$p_I(0) = P\left[\{\zeta : \zeta \in A^c\}\right] = 1 - p$$
$$p_I(1) = P[\{\zeta : \zeta \in A\}] = p$$

We call I_A the **Bernoulli random variable**. Note that $p_I(1) + p_I(2) = 1$.

Example 5: Message Transmissions

Let X be the number of times a message needs to be transmitted until it arrives correctly at its destination. Find the pmf of X. Find the probability that X is an even number.

X is a discrete random variable taking on values from $S_X = \{1, 2, 3, \ldots\}$. The event $\{X = k\}$ occurs if the underlying experiment finds $k-1$ consecutive erroneous transmissions ("failures") followed by a error-free one ("success"):

$$p_X(k) = P[X = k] = P[00\ldots01] = (1-p)^{k-1}p = q^{k-1}p \quad k = 1, 2, \ldots$$

We call X the **geometric random variable**, and we say that X is geometrically distributed. In **Eq. (2.42b)**, we saw that the sum of the geometric probabilities is 1 .

$$P[X \text{ is even }] = \sum_{k=1}^{\infty} p_X(2k) = p \sum_{k=1}^{\infty} q^{2k-1} = p\frac{1}{1-q^2} = \frac{1}{1+q}$$

Example 6: Transmission Errors

A binary communications channel introduces a bit error in a transmission with probability p. Let X be the number of errors in n independent transmissions. Find the pmf of X. Find the probability of one or fewer errors.

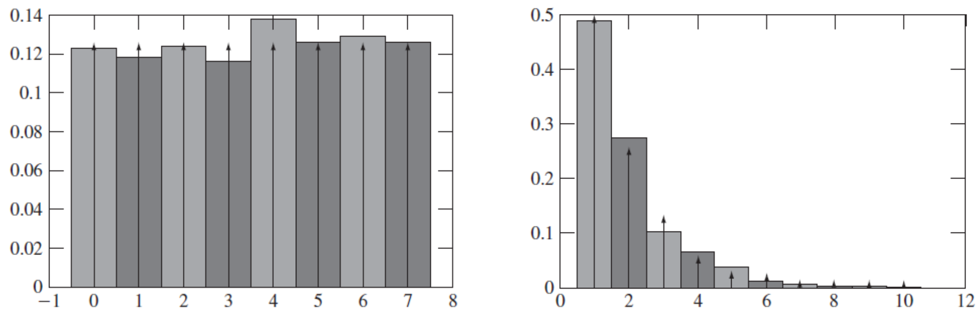
X takes on values in the set $S_X = \{0, 1, \ldots, n\}$. Each transmission results in a "0" if there is no error and a "1" if there is an error, $P[\text{ "1"}] = p$ and $P[\text{ "0"}] = 1 - p$. The probability of k errors in n bit transmissions is given by the probability of an error pattern that has k1's and $n-k$0's:

$$p_X(k) = P[X = k] = \binom{n}{k} p^k (1-p)^{n-k} \quad k = 0, 1, \ldots, n$$

We call X the **binomial random variable**, with parameters n and p. In **Eq. (2.39b)**, we saw that the sum of the binomial probabilities is 1.

$$P[X \leq 1] = \binom{n}{0} p^0 (1-p)^{n-0} + \binom{n}{1} p^1 (1-p)^{n-1} = (1-p)^n + np(1-p)^{n-1}$$

Finally, let's consider the relationship between relative frequencies and the pmf $p_X(x_k)$. Suppose we perform n independent repetitions to obtain n observations of the discrete random variable X. Let $N_k(n)$ be the number of times the event $X = x_k$ occurs and let $f_k(n) = N_k(n)/n$ be the corresponding relative frequency. As n becomes large we expect that $f_k(n) \to p_X(x_k)$. Therefore the graph of relative frequencies should approach the graph of the pmf. Figure 3a(a) shows the graph of relative frequencies for 1000 repetitions of an experiment that generates a uniform random variable from the set $\{0, 1, \ldots, 7\}$ and the corresponding pmf. Figure 3b(b) shows the graph of relative frequencies and pmf for a geometric random variable with $p = 1/2$ and $n = 1000$ repetitions. In both cases we see that the graph of relative frequencies approaches that of the pmf.



(a) Events A and B are independent. **(b)** Events A and C are not independent.

Figure 3

2 The Probability Density Function

The **probability density function** of X (pdf) is defined as the derivative of $F_X(x)$ at points where the cdf is differentiable:

$$f_X(x) = \frac{dF_X(x)}{dx} \tag{5}$$

- The pdf is an alternative, and more useful, way of specifying the information contained in the cumulative distribution function.

- We can imagine dividing the real line a sequence of disjoint intervals labeled by integers to create a discrete random variable with pmf given by the probability of X falling in each interval. We will soon see that the pdf and pmf have similar properties.

The pdf represents the "density" of probability at the point x in the following sense: The probability that X is in a small interval in the vicinity of x, that is $\{x < X \leq x + h\}$, is

$$\begin{aligned} P[x < X \leq x + h] &= F_X(x + h) - F_X(x) \\ &= \frac{F_X(x + h) - F_X(x)}{h} h. \end{aligned}$$

If the cdf has a derivative at x, then as h becomes very small

$$P[x < X \leq x + h] \simeq f_X(x)h \tag{6}$$

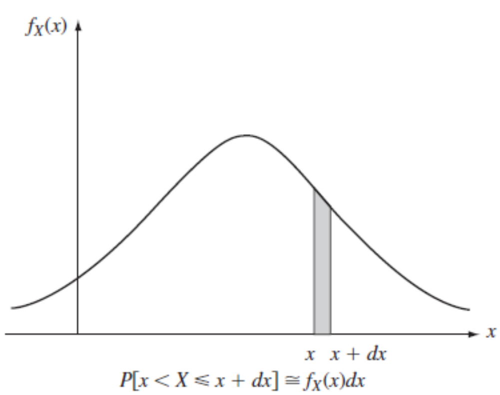
Thus $f_X(x)$ represents the "density" of probability at the point x in the sense that the probability that X is in a small interval in the vicinity of x is approximately $f_X(x)h$. The derivative of the cdf, when it exists, is positive since the cdf is a nondecreasing function of x, thus

$$\text{(i)} \quad f_X(x) \geq 0. \tag{7}$$

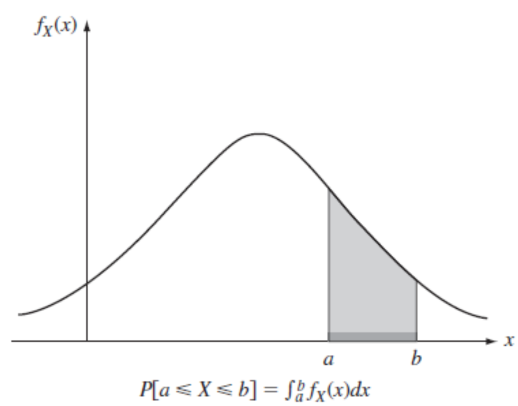
- Equations (6) and (7) provide an alternative approach to specifying the probabilities involving the random variable X. We begin with a pdf $f_X(x)$ which specifies the probabilities of events of the form "X falls in a small interval of width dx about the point x," as shown in Fig. 4(a).

- The probability of an event involving X is then obtained by adding the probabilities of intervals of width dx. As the widths of the intervals approach zero, we obtain an integral in terms of the pdf. For example, the probability of an interval $[a, b]$, is

$$\text{(ii)} \quad P[a \leq X \leq b] = \int_a^b f_X(x)dx \tag{8}$$



(a) The pdf specifies the probability of intervals of infinitesimal width.



(b) The probability of an interval $[a, b]$ is the area under the pdf in that interval.

Figure 4

The probability of an interval is therefore the area under $f_X(x)$ in that interval as shown in Fig. 4(b). *The probability of any event that consists of* *the union of disjoint intervals can thus be found by adding the integrals of the pdf over each of the intervals.*

The cdf of X can be obtained by integrating the pdf:

$$\text{(iii)} \quad F_X(x) = \int_{-\infty}^{x} f_X(t)dt. \tag{9}$$

In **Section 3.2**, we defined a continuous random variable as a random variable X whose cdf was given by Eq. (9). Since the probabilities of all events involving X can be written in terms of the cdf, it then follows that these probabilities can be written in terms of the pdf. Thus the pdf completely specifies the behavior of continuous random variables.

By letting x tend to infinity in Eq. (9), we obtain a normalization condition for pdf's:

$$\text{(iv)} \quad 1 = \int_{-\infty}^{+\infty} f_X(t)dt \tag{10}$$

The pdf reinforces the intuitive notion of probability as having attributes similar to "physical mass." Thus Eq. (8) states that the probability "mass" in an interval is the integral of the "density of probability mass" over the interval. Equation (10) states that the total mass available is one unit.

A valid pdf can be formed from any nonnegative, piecewise continuous function $g(x)$ that has a finite integral:

$$\int_{-\infty}^{\infty} g(x)dx = c < \infty$$

By letting $f_X(x) = g(x)/c$, we obtain a function that satisfies the normalization condition. Note that the pdf must be defined for all real values of x; if X does not take on values from some region of the real line, we simply set $f_X(x) = 0$ in the region.

Example 7: Uniform Random Variable

The pdf of the uniform random variable is given by:

$$f_X(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & x < a \text{ and } x > b \end{cases}$$

and is shown in Fig. 5(a). The cdf is found from Eq. (9):

$$F_X(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \le x \le b \\ 1 & x > b \end{cases}$$

The cdf is shown in Fig. 5(b)

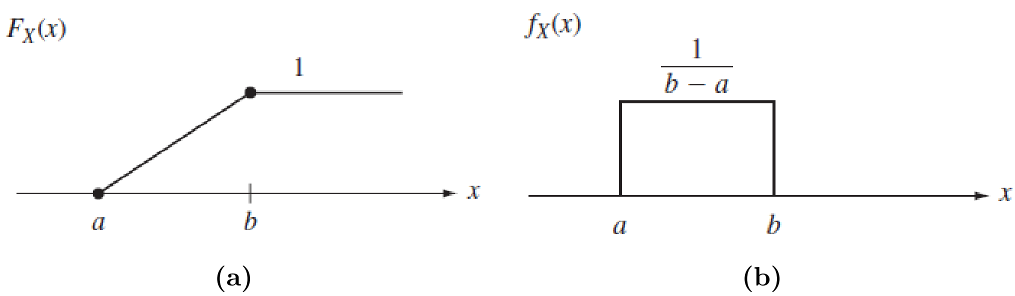


Figure 5: cdf (a) and pdf (b) of a continuous random variable

Example 8: Exponential Random Variable

The transmission time X of messages in a communication system obey the exponential probability law with parameter λ, that is,

$$P[X > x] = e^{-\lambda x} \quad x > 0.$$

Find the cdf of X.

The cdf of X is $F_X(x) = P[X \le x] = 1 - P[X > x]$:

$$F_X(x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-\lambda x} & x \ge 0 \end{cases} \tag{11}$$

The pdf is obtained by applying Eq. (5):

$$f_X(x) = F'_X(x) = \begin{cases} 0 & x < 0 \\ \lambda e^{-\lambda x} & x \ge 0 \end{cases}.$$

Example 9: Laplacian Random Variable

The pdf of the samples of the amplitude of speech waveforms is found to decay exponentially at a rate α, so the following pdf is proposed:

$$f_X(x) = ce^{-\alpha|x|} \quad -\infty < x < \infty.$$

Find the constant c, and then find the probability $P[|X| < v]$.

We use the normalization condition in (iv) to find c:

$$1 = \int_{-\infty}^{\infty} ce^{-\alpha|x|}dx = 2\int_{0}^{\infty} ce^{-\alpha x}dx = \frac{2c}{\alpha}$$

Therefore $c = \alpha/2$. The probability $P[|X| < v]$ is found by integrating the pdf:

$$\begin{aligned} P[|X| < v] &= \frac{\alpha}{2}\int_{-v}^{v} e^{-\alpha|x|}dx = 2\left(\frac{\alpha}{2}\right)\int_{0}^{\nu} e^{-\alpha x}dx \\ &= 1 - e^{-\alpha v}. \end{aligned} \tag{12}$$

The pdf in Eq. (12) is called the **Laplacian pdf**.

2.1 pdf of Discrete Random Variables

The derivative of the cdf does not exist at points where the cdf is not continuous. Thus the notion of pdf as defined by Eq. (5) does not apply to discrete random variables at the points where the cdf is discontinuous.
We can generalize the definition of the probability density function by noting the relation between the unit step function and the delta function. The **unit step function** is defined as

$$u(x) = \begin{cases} 0 & x < 0 \\ 1 & x \geq 0 \end{cases}$$

The **delta function** $\delta(t)$ is related to the unit step function by the following equation:

$$u(x) = \int_{-\infty}^{x} \delta(t)dt$$

- Basically we can treat the unit step function as the integral of a delta function at the origin, and we can view the delta function as the "derivative" of the unit step function.

A translated unit step function is then the integral of a delta function at the point $x = x_0$:

$$u(x - x_0) = \int_{-\infty}^{x-x_0} \delta(t)dt = \int_{-\infty}^{x} \delta(t' - x_0)dt'. \tag{13}$$

The cdf of a discrete random variable can be expressed as a sum of unit step functions with amplitudes given by the pmf:

$$F_X(x) = \sum_k p_X(x_k)\, u(x - x_k) \tag{14}$$

where the probability mass function is $p_X(x_k) = P[X = x_k]$. We can generalize the definition of the pdf $f_X(x)$ by taking the derivative of the above cdf:

$$f_X(x) = \sum_k p_X(x_k)\, \delta(x - x_k) \tag{15}$$

Substitution of Eq. (15) into Eq. (9) then yields (14) as required. Thus the generalized definition of pdf places a delta function of weight $P[X = x_k]$ at the points x_k where the cdf is discontinuous.

2.2 Conditional cdf's and pdf's

Conditional cdf's can be defined in a straightforward manner by replacing the probability in Eq. (3.1) by a conditional probability. For example, if some event A concerning X is given, then the **conditional cdf of X given A** is defined by

$$F_X(x \mid A) = \frac{P[\{X \leq x\} \cap A]}{P[A]} \quad \text{if } P[A] > 0$$

The **conditional pdf of X given A** is then defined by

$$f_X(x \mid A) = \frac{d}{dx} F_X(x \mid A)$$

Example 10

The lifetime X of a machine has a continuous cdf $F_X(x)$. Find the conditional cdf and pdf given the event $A = \{X > t\}$ (i.e., "the machine is still working at time t''.")

The conditional cdf is

$$\begin{aligned} F_X(x \mid X > t) &= P[X \leq x \mid X > t] \\ &= \frac{P[\{X \leq x\} \cap \{X > t\}}{P[X > t]} \end{aligned}$$

The intersection of the two events in the numerator is equal to the empty set when $x < t$ and to $\{t < X \leq x\}$ when $x \geq t$. Thus

$$F_X(x \mid X > t) = \begin{cases} 0 & x < t \\ \frac{F_X(x) - F_X(t)}{1 - F_X(t)} & x \geq t \end{cases}$$

The conditional pdf is found by differentiating with respect to x :

$$f_X(x \mid X > t) = \frac{f_X(x)}{1 - F_X(t)} \quad x \geq t$$

3 The Expected Value of Random Variables

The cdf or pdf are functions that completely describe the behavior of a random variable. In some situations we are interested in a few parameters that summarize the information provided by these functions.

- Fig. 6 shows the results of many repetitions of an experiment that produces two random variables. The random variable X varies about the value 0 whereas the random variable Y varies about the value 5 . We can also see that Y is more spread out than X.

- In this section we introduce parameters that quantify these properties that help distinguish the behavior of different random variables.

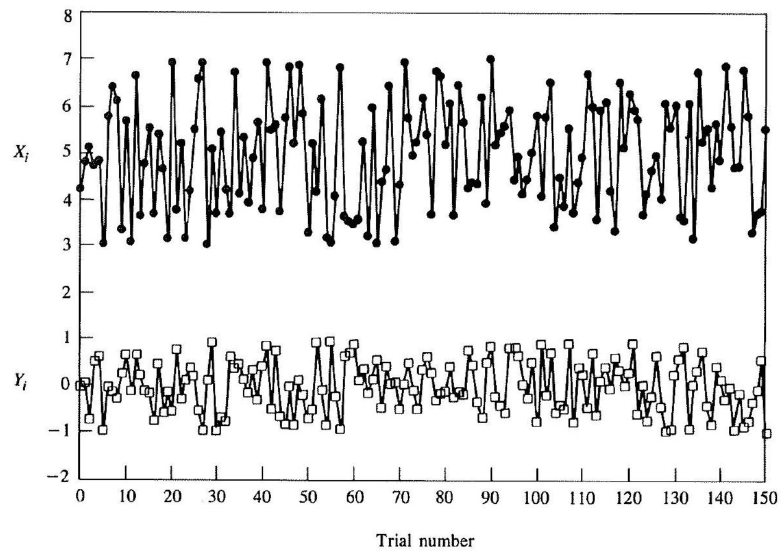


Figure 6: The graphs show 150 repetitions of the experiments yielding the random variable X and the random variable Y. It is clear that X takes on values centered about the value 5 while Y takes on values centered about 0 . It is also clear that X is more spread out than Y.

3.1 The Expected Value of X

The **expected value** or **mean** of a random variable X is defined by

$$E[X] = \int_{-\infty}^{+\infty} t f_X(t) dt. \tag{16}$$

If X is a discrete random variable, substitution of Eq. (15) into Eq. (16) yields

$$E[X] = \sum_k x_k p_X(x_k) \tag{17}$$

The expected value $E[X]$ is defined if the above integral or sum converge absolutely, that is,

$$E[|X|] = \int_{-\infty}^{+\infty} |t| f_X(t) dt \quad < \infty$$

or

$$E[|X|] = \sum_k |x_k| \, p_X(x_k) \quad < \infty$$

- The mean is simply the sum or integral of the values that X can take on weighted by their corresponding probability or probability density.
- There are random variables for which the above expressions do not converge, for example the Zipf, Pareto, and Cauchy random variables introduced later.
- If we view $f_X(x)$ as the distribution of mass on the real line, then $E[X]$ represents the center-of-mass of this distribution.

The expression for the mean of X appears naturally if we use relative frequencies. Consider a discrete finite random variable X and let x_k be a possible outcome. Suppose that we repeat the experiment n times and we observe the values $X(1), X(2), ..., X(n)$. Consider the average of the observed values, the so-called **sample mean**:

$$\begin{aligned} \langle X \rangle_n &= \frac{1}{n} \sum_{j=1}^{n} X(j) & (18) \\ &= \frac{1}{n} \sum_{k=0}^{\infty} x_k N_k(n) & (19) \end{aligned}$$

The first expression adds the number of observed values in the n trials and divides the sum by n. The second expression counts how many of these observations had value x_k, and then computes the total and divides by n.[2] As n gets large, the ratio $N_k(n)/n$ in the second expression approaches p_k. Thus the average number of the observed values approaches

$$\langle X \rangle_n \to \sum_{k=0}^{\infty} x_k p_k \triangleq E[X] \tag{20}$$

The expression on the right-hand side is the **expected value of** X and is completely determined by the pmf.

[2]Suppose you pull out the following change from your pocket: 1 quarter, 1 dime, 1 quarter, 1 nickel. Equation (1.7) says your total is $25 + 10 + 25 + 5 = 65$ cents. Equation (1.8) says your total is $(1)5 + (1)10 + (2)(25) = 65$ cents.

Example 11: Mean of a Uniform Random Variable

The mean for a uniform random variable is given by

$$E[X] = (b-a)^{-1} \int_a^b t dt = \frac{a+b}{2}$$

which is exactly the midpoint of the interval $[a, b]$. The results shown in Fig. 6 were obtained by repeating experiments in which outcomes were random variables X and Y that had uniform cdf's in the intervals $[-1, 1]$ and $[3, 7]$, respectively. The respective expected values, 0 and 5 , correspond to the values about which X and Y tend to vary.

The result in Example 11 could have been found immediately by noting that $E[X] = m$ when the pdf is symmetric about a point m. That is, if

$$f_X(m-x) = f_X(m+x) \quad \text{for all } x$$

then, assuming that the mean exists,

$$0 = \int_{-\infty}^{+\infty} (m-t) f_X(t) dt = m - \int_{-\infty}^{+\infty} t f_X(t) dt$$

The first equality above follows from the symmetry of $f_X(t)$ about $t = m$ and the odd symmetry of ($m - t$) about the same point. We then have that $E[X] = m$.

The following expressions are useful when X is a nonnegative random variable:

$$E[X] = \int_0^{\infty} (1 - F_X(t))\, dt \quad \text{if } X \text{ continuous and nonnegative} \tag{21}$$

and

$$E[X] = \sum_{k=0}^{\infty} P[X > k] \quad \text{if } X \text{ nonnegative, integer-valued.} \tag{22}$$

The derivation of these formulas is discussed in the problems.

Example 12: Mean of Exponential Random Variable

The time X between customer arrivals at a service station has an exponential pdf with parameter λ. Find the mean interarrival time.

Substituting Eq. (11) into Eq. (16) we obtain

$$E[X] = \int_0^\infty t\lambda e^{-\lambda t} dt$$

We evaluate the integral using integration by parts $\left(\int u dv = uv - \int v du\right)$, with $u = t$ and $dv = \lambda e^{-\lambda t} dt$:

$$\begin{aligned} E[X] &= \lambda t e^{-\lambda t}\Big|_0^\infty - \int_0^\infty e^{-\lambda t} dt \\ &= \lim_{t\to\infty} t e^{-\lambda t} - 0 - \left\{\frac{-e^{-\lambda t}}{\lambda}\right\}_0^\infty \\ &= \lim_{t\to\infty} \frac{e^{-\lambda t}}{\lambda} + \frac{1}{\lambda} = \frac{1}{\lambda} \end{aligned}$$

where we have used the fact that $e^{-\lambda t}$ and $te^{-\lambda t}$ go to zero as t approaches infinity.

For this example, Eq. (21) is much easier to evaluate:

$$E[X] = \int_0^\infty e^{-\lambda t} dt = \frac{1}{\lambda}$$

Example 13

Let N be the number of times a computer polls a terminal until the terminal has a message ready for transmission. If we suppose that the terminal produces messages according to a sequence of independent Bernoulli trials, then N has a geometric distribution. Find the mean of N.

The expected value of the geometric random variable using Eq. (17) is

$$E[N] = \sum_{k=1}^{\infty} kpq^{k-1}$$

This expression is readily evaluated by differentiating the series

$$\frac{1}{1-x} = \sum_{k=0}^{\infty} x^k$$

to obtain

$$\frac{1}{(1-x)^2} = \sum_{k=0}^{\infty} kx^{k-1}$$

Letting $x = q$, we obtain

$$E[N] = p\frac{1}{(1-q)^2} = \frac{1}{p}$$

The direct evaluation of Eq. (22) is easy in this case:

$$E[N] = \sum_{k=0}^{\infty} P[N > k] = \sum_{k=0}^{\infty} q^k = \frac{1}{1-q}$$

where we have used the fact that $P[N > k] = q^k$ for $k = 0, 1, 2, \ldots$.

Suppose that $q = .6$ in the above example. Then $E[N] = 2.5$, which is not a value that can be assumed by N. Thus the statement " N equals 2.5 on the average" does not make sense. (This reminds us of reading in the newspaper that the "average household has 3.5 persons.") What does make sense is the statement that the arithmetic average of a large number of repetitions of an experiment will be close to 2.5 .

3.1.1 The Expected Value of $Y = g(X)$

Suppose that we are interested in finding the expected value of $Y = g(X)$. The direct approach involves first finding the pdf of Y, and then the evaluation of $E[Y]$ using Eq. (16). We now show that $E[Y]$ can be found directly in terms of the pdf of X :

$$E[Y] = \int_{-\infty}^{\infty} g(x) f_X(x) dx \tag{23}$$

To see how Eq. (23) comes about, suppose that we divide the y-axis into intervals of length h, we index the intervals with the index k and we let y_k be the value in the center of the kth interval. The expected value of Y is approximated by the following sum:

$$E[Y] \simeq \sum_k y_k f_Y(y_k)\, h$$

Suppose that $g(x)$ is strictly increasing, then the k th interval in the y-axis has a unique corresponding equivalent event of width h_k in the x-axis as shown in Fig. 7. Let x_k be the value in the k th interval such that $g(x_k) = y_k$, then

$$E[Y] \simeq \sum_k g(x_k) f_X(x_k) h_k$$

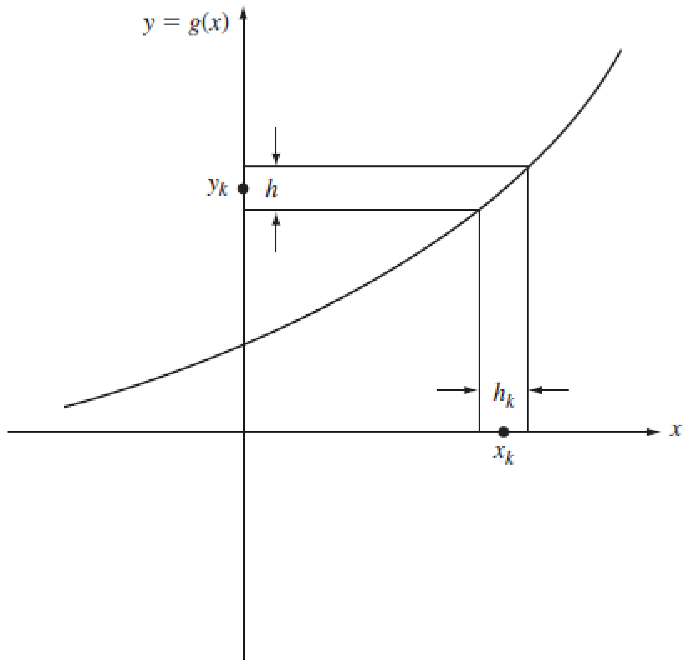


Figure 7: Two infinitesimal equivalent events.

By letting h approach zero, we obtain Eq. (23). This equation is valid even if $g(x)$ is not strictly increasing. The derivation for the general case involves taking into account the fact that each y-interval has an equivalent event as in Figure 8.

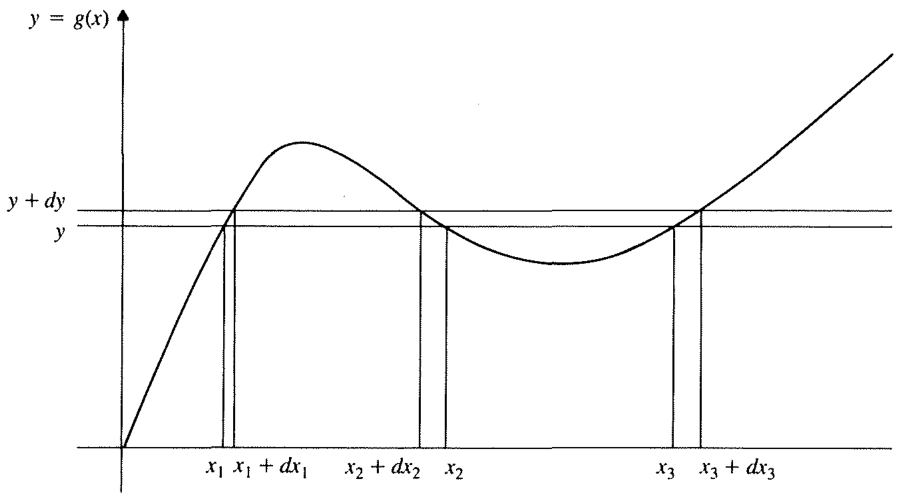


Figure 8: The equivalent event of $\{y < Y \leq y + dy\}$ is $\{x_1 < X \leq x_1 + dx_1\} \cup \{x_2 < X \leq x_2\} \cup \{x_3 < X \leq x_3\}$

Example 14

Let $Y = g(X) = A\cos(X)$, where A is a constant and X is uniformly distributed in the interval $(0, 2\pi]$. The expected value of Y is then

$$E[Y] = \int_0^{2\pi} A\cos(x)\frac{1}{2\pi}dx = -\,A\sin(x)|_0^{2\pi} = 0$$

Example 15: Expected Values of the Indicator Function

Let $g(X) = I_C(X)$ be the indicator function for the event $\{X \text{ in } C\}$, where C is some interval or union of intervals in the real line:

$$g(X) = \begin{cases} 0 & X \text{ not in } C \\ 1 & X \text{ in } C \end{cases}$$

then

$$E[Y] = \int_{-\infty}^{+\infty} g(X)f_X(x)dx = \int_C f_X(x)dx = P[X \text{ in } C]$$

Thus the expected value of the indicator of an event is equal to the probability of the event.

The following two simple, but useful, properties follow immediately from Eq. (23). Let c be some constant, then

$$E[c] = \int_{-\infty}^{\infty} c f_X(x) dx = c \int_{-\infty}^{\infty} f_X(x) dx = c \tag{24}$$

and

$$E[cX] = \int_{-\infty}^{\infty} cx f_X(x) dx = c \int_{-\infty}^{\infty} x f_X(x) dx = cE[X] \tag{25}$$

The expected value of a sum of functions of a random variable is equal to the sum of the expected values of the individual functions:

$$\begin{aligned} E[Y] &= E\left[\sum_{k=1}^{n} g_k(X)\right] \\ &= \int_{-\infty}^{\infty} \sum_{k=1}^{n} g_k(x) f_X(x) dx = \sum_{k=1}^{n} \int_{-\infty}^{\infty} g_k(x) f_X(x) dx \\ &= \sum_{k=1}^{n} E\left[g_k(X)\right] \end{aligned} \tag{26}$$

Example 16

Let $Y = g(X) = a_0 + a_1 X + a_2 X^2 + \cdots + a_n X^n$, where a_k are constants, then

$$\begin{aligned} E[Y] &= E\left[a_0\right] + E\left[a_1 X\right] + \cdots + E\left[a_n X^n\right] \\ &= a_0 + a_1 E[X] + a_2 E\left[X^2\right] + \cdots + a_n E\left[X^n\right] \end{aligned}$$

where we have used Eq. (26), and Eqs. (24) and (25). A special case of this result is that

$$E[X + c] = E[X] + c$$

that is, we can shift the mean of a random variable by adding a constant to it.

4 Variance of X

- The expected value $E[X]$, by itself, provides us with very limited information about X. For example, if we know that $E[X] = 0$, then it could be that X is zero all the time. However it could be as well that X is equally likely to take on extremely large positive and negative values.

- We are therefore interested not only in the mean of a random variable, but also in the extent of the random variable's variation about its mean.

- Let the deviation of X about its mean be $D = X - E[X]$, then D can take on positive and negative values. Since we are interested in the magnitude of the variations only, it is convenient to work with D^2, which is always positive.

- The **variance** of the random variable X is defined as the mean squared variation $E\left[D^2\right]$

$$\mathrm{VAR}[X] = E\left[(X - E[X])^2\right] \tag{27}$$

By taking the square root of the variance we obtain a quantity with the same units as X. The **standard deviation** of the random variable X is defined by

$$\mathrm{STD}[X] = \mathrm{VAR}[X]^{1/2}$$

The $\mathrm{STD}[X]$ is used as a measure of the "width" or "spread" of a distribution.

The expression in Eq. (27) can be simplified as follows:

$$\begin{aligned}
\mathrm{VAR}[X] &= E\left[\left(X^2 - 2E[X]X + E[X]^2\right]\right. \\
&= E\left[X^2\right] - 2E[X]E[X] + E[X]^2 \\
&= E\left[X^2\right] - E[X]^2
\end{aligned} \tag{28}$$

where the fact that $E[X]$ is a constant and Eqs. (24) and (25) have been used.

Example 17: Variance of Uniform Random Variable

Find the variance the random variable X that is uniformly distributed in the interval $[a, b]$.

Since the mean of X is $(a + b)/2$,

$$\text{VAR}[X] = \frac{1}{b-a} \int_a^b \left(x - \frac{a+b}{2}\right)^2 dx$$

Let $y = (x - (a + b)/2)$,

$$\text{VAR}[X] = \frac{1}{b-a} \int_{(b-a)/2}^{(b-a)/2} y^2 dy = \frac{(b-a)^2}{12}$$

The random variables in Fig. 6 were uniformly distributed in the interval $[-1, 1]$ and $[3, 7]$, respectively. Their variances are then $1/3$ and $4/3$. The corresponding standard deviations are 0.577 and 1.155 .

Example 18: Variance of Geometric Random Variable

Find the variance of the geometric random variable.

Differentiate the term $(1 - x)^{-2}$ in Example 13 to obtain

$$\frac{2}{(1-x)^3} = \sum_{k=0}^{\infty} k(k-1)x^{k-2}.$$

Letting $x = q$ and multiplying both sides by pq, we obtain

$$\frac{2q}{(1-q)^2} = E\left[N^2\right] - E[N]$$

Recalling that $E[N] = 1/p$, we find that $E\left[N^2\right] = (1 + q)/p^2$. The variance is then obtained using Eq. (28).

$$\text{VAR}[N] = E\left[N^2\right] - E[N]^2 = \frac{q}{p^2}.$$

The following properties are easy to show.

$$\text{VAR}[c] = 0$$

$$\text{VAR}[X + c] = \text{VAR}[X]$$

$$\text{VAR}[cX] = c^2\,\text{VAR}[X],$$

where c is a constant.

The mean and variance are the two most important parameters used in summarizing the pdf of a random variable. Other parameters are occasionally used. For example, the *skewness* defined by $E[(X - E[X])^3]\,/\,\text{STD}[X]^3$ measures the degree of asymmetry about the mean. It is easy to show that if a pdf is symmetric about its mean, then its skewness is zero. The point to note with these parameters of the pdf is that each involves the expected value of a higher power of X. Indeed we show in a later section that, under certain conditions, a pdf is completely specified if the expected value of all the powers of X are known. These expected values are called the **moments** of X.

The nth **moment of the random variable** X is defined by

$$E\left[X^n\right] = \int_{-\infty}^{\infty} x^n f_X(x)dx$$

The mean and variance can be seen to be defined in terms of the first two moments, $E[X]$ and $E\left[X^2\right]$.

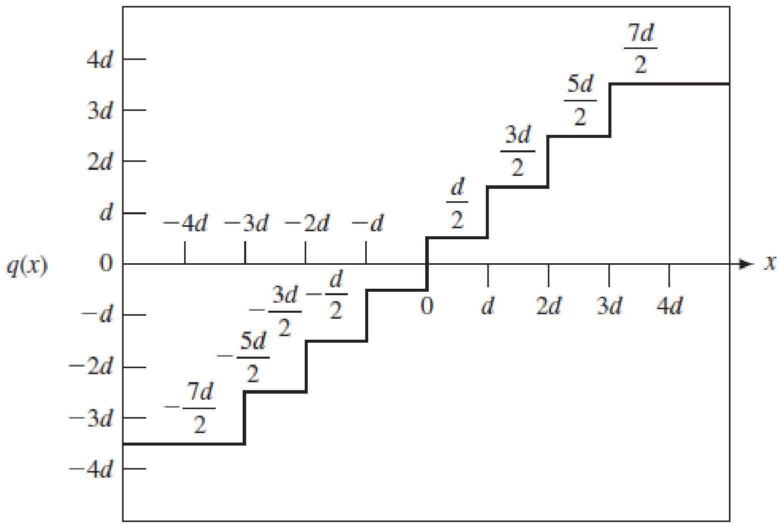
Example 19: Analog-to-Digital Conversion: A Detailed Example

A quantizer is used to convert an analog sigmal (e.g., speech or audio) into digital form. A quantizer maps a random voltage X into the nearest point $q(X)$ from a set of 2^R representation values, as shown in Fig. 9a(a). The value X is then approximated by $q(X)$, which is identified by an R bit binary number. In this manner, an "analog" voltage X, that can assume a continuum of values, is converted into an R bit number.

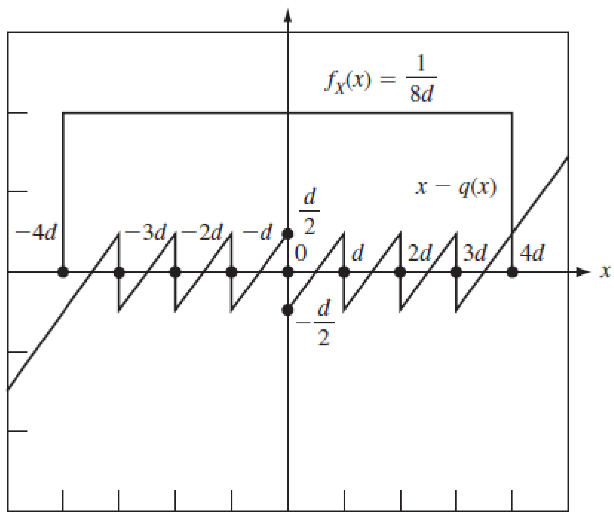
The quantizer introduces an error $Z = X - q(X)$ as shown in Fig. 9b(b). Note that Z is a function of X and that it ranges in value between $-d/2$ and $d/2$, where d is the quantizer step size. Suppose that X has a uniform distribution in the interval $[-x_{\max}, x_{\max}]$, that the quantizer has 2^R levels, and that $2x_{\max} = 2^R d$. It is easy to show that Z is uniformly distributed in the interval $[-d/2, d/2]$.

From our previous discussion on the mean of a uniform random variable we have,

$$E[Z] = \frac{d/2 - d/2}{2} = 0$$



(a) A uniform quantizer maps the input x into the closest point form the set $\{\pm d/2, \pm 3d/2, \pm 5d/2, \pm 7d/2\}$



(b) The uniform quantizer error for the input x is $x - q(x)$

Figure 9

The error Z thus has mean zero.

By Example 17,

$$\mathrm{VAR}[Z] = \frac{(d/2 - (-d/2))^2}{12} = \frac{d^2}{12}.$$

This result is approximately correct for any pdf that is approximately flat over each quantizer interval. This is the case when 2^R is large.

The approximation $q(x)$ can be viewed as a "noisy" version of X since

$$Q(X) = X - Z$$

where Z is the quantization error Z. The measure of goodness of a quantizer is specified by the SNR ratio, which is defined as the ratio of the variance of the "signal" X to the variance of the distortion or "noise" Z :

$$\begin{aligned}\mathrm{SNR} &= \frac{\mathrm{VAR}[X]}{\mathrm{VAR}[Z]} = \frac{\mathrm{VAR}[X]}{d^2/12} \\ &= \frac{\mathrm{VAR}[X]}{x_{\max}^2/3} 2^{2R}\end{aligned}$$

where we have used the fact that $d = 2x_{\max}/2^R$. When X is nonuniform, the value $x_{\max}$ is selected so that $P\left[|X| > x_{\max}\right]$ is small. A typical choice is $x_{\max} = 4\mathrm{STD}[X]$. The SNR is then

$$\mathrm{SNR} = \frac{3}{16} 2^{2R}$$

This important formula is often quoted in decibels:

$$\mathrm{SNRdB} = 10 \log_{10} \mathrm{SNR} = 6R - 7.3 \text{ dB}$$

The SNR increases by a factor of 4(6 dB) with each additional bit used to represent X. This makes sense since each additional bit doubles the number of quantizer levels, which in turn reduces the step size by a factor of 2. The variance of the error should then be reduced by the square of this, namely $2^2 = 4$.