Weekly Notes for EE2012 2014/15 – Week 9

T. J. Lim

March 19, 2015

Book sections covered this week: 4.5, 5.1

1 Functions of a Random Variable

1.1 Problem Statement

Random variables appear in engineering problems in many guises, e.g. as a time interval, as a count of the number of occurrences of some phenomenon, as a physical measurement such as weight or voltage. It is highly likely that functions of these random variables represent useful things, such as area being proportional to the square of a length, or time in minutes being one-sixtieth of time in seconds. Therefore, we are interested in *functions of a random variable*, $Y = g(X)$, where $g(x)$ is a function with a domain[1] $\mathcal{D}$ that is a subset of the range of X, or $\mathcal{D} \subset S_X$.

The question of interest is: If we know the distribution of X (i.e. its PMF, PDF or CDF), how do we find the distribution of $Y = g(X)$?

1.2 General Procedure

The first thing to do is to identify the sample space of Y. This is usually straightforward, once we know the sample space of X and the function $g(x)$. For instance,

$$\begin{aligned} S_X = \{0, 1, 2, \ldots\}, g(x) = x^2 &\quad\Rightarrow\quad S_Y = \{0^2, 1^2, 2^2, \ldots,\} \\ S_X = \mathbb{R}, g(x) = \mathrm{sgn}(x) &\quad\Rightarrow\quad S_Y = \{-1, +1\} \\ S_X = \mathbb{R}_+, g(x) = \ln(x) &\quad\Rightarrow\quad S_Y = \mathbb{R}. \end{aligned}$$

If S_Y is countable, i.e. Y is a discrete random variable, then we will find the PMF of Y, as described in the next section. If Y is mixed or continuous, then we can find the CDF of Y. If Y is continuous, we also have the option of finding the PDF directly from the PDF of X.

[1] i.e. the set of allowable inputs to the function.

1.3 Finding the PMF of Discrete Y

Suppose we know that Y is discrete. Then the PMF of Y is

$$p_Y(y_k) = P[Y = y_k] \tag{1}$$

where $y_k \in S_Y$. The event $\{Y = y_k\}$ is equivalent to $\{g(X) = y_k\}$, in turn equivalent to $X \in A_k$ where $A_k = \{x : g(x) = y_k\}$. Therefore, $p_Y(y_k) = P[X \in A_k]$, which simplifies to

$$p_Y(y_k) = \sum_{x \in A_k} p_X(x), \quad \text{if } X \text{ is discrete} \tag{2}$$

or for continuous X,

$$p_Y(y_k) = \int_{A_k} f_X(x)dx. \tag{3}$$

In principle $P[X \in A_k]$ can be found since we know the distribution of X, and hence we can obtain the PMF of Y.

Example 1 (Discrete X): If $X \sim \mathcal{B}(4, 0.2)$, what is the PMF of $Y = 2X + 1$? How about the PMF of $Z = |X - 2|$?

Ans: We can write out the following table of x, y and z values, where $x \in \{0, 1, 2, 3, 4\}$, $y = 2x + 1$ and $z = |x - 2|$.

x	0	1	2	3	4
y	1	3	5	7	9
z	2	1	0	1	2

From this table, we see that for these values of x, y can only take values in $\{1, 3, 5, 7, 9\}$, which is therefore the sample space or range of Y. Also, the event $\{Y = 1\}$ is equivalent to $\{X = 0\}$, and $\{Y = 3\}$ is equivalent to $\{X = 1\}$, etc. and so

$$p_Y(1) = p_X(0), \;\; p_Y(3) = p_X(1), \text{ etc.}$$

In general, $p_Y(2k + 1) = p_X(k)$, $k = 0, 1, 2, 3, 4$.

For $Z = |X-2|$, the event $\{Z = 2\}$ is equivalent to $\{X = 0\}$ or $\{X = 4\}$, because either of the latter two events must occur for the former to occur. Therefore, $p_Z(2) = p_X(0) + p_X(4)$. Similarly, $p_Z(1) = p_X(1) + p_X(3)$. Finally, $\{Z = 2\}$ is equivalent to $\{X = 2\}$, and hence $p_Z(2) = p_X(2)$. ■

Example 2 (Continuous X): Let X be a continuous random variable, and $g(x) = \text{sgn}(x)$. Then $S_Y = \{-1, +1\}$, and

$$p_Y(-1) \;=\; P[Y = -1] = P[X < 0] = \int_{-\infty}^{0} f_X(t)dt \tag{4}$$

$$p_Y(+1) \;=\; P[Y = +1] = P[X \geq 0] = \int_{0}^{\infty} f_X(t)dt. \tag{5}$$

So if for example $f_X(x) = e^{-(x+1)}$, $x > -1$, then

$$p_Y(-1) = \int_{-1}^{0} e^{-(x+1)} dx = 1 - e^{-1},$$

and $p_Y(1) = e^{-1}$. ■

Example 3: Let X be binomial with parameters n and p, and $Y = X^2$. Then it should be clear that

$$S_Y = \{k^2 : k = 0, 1, \ldots, n\}$$

and that $\{Y = k^2\} = \{X = k\}$. Therefore, Y has the PMF

$$p_Y(k^2) = p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \ldots, n.$$

Example 4: Let X be binomial with $n = 6$ and probability of success p, and $Y = |X - 3|$. Then $S_X = \{0, 1, 2, 3, 4, 5, 6\}$, and $S_Y = \{0, 1, 2, 3\}$. The event $\{Y = 0\} = \{X = 3\}$, while $\{Y = k\} = \{X = 3 + k\} \cup \{X = 3 - k\}$ for $k = 1, 2, 3$. Therefore, the PMF of Y is

$$p_Y(k) = \begin{cases} \binom{6}{3} p^3 (1-p)^3 & k = 0 \\ \binom{6}{3-k} p^{3-k} (1-p)^{3+k} + \binom{6}{3+k} p^{3+k} (1-p)^{3-k} & k = 1, 2, 3. \end{cases} \tag{6}$$

The above examples illustrate the general rule that Y is discrete if and only if

- X is discrete, or
- $g(x)$ is a staircase function, i.e. $g(x) = a_0 + \sum_k a_k u(x - x_k)$,

or both, and also that the distribution of $Y = g(X)$ is usually not too hard to derive in this case.

1.4 Finding the CDF of Y

If Y is not discrete, then we can always try to find its CDF from first principles, i.e.

$$F_Y(y) = P[Y \leq y] = P[g(X) \leq y]. \tag{7}$$

Since the event $\{g(X) \leq y\}$ involves only X, its probability can in principle be found, because the distribution of X is known. In practice, it can of course be rather challenging and often actually impossible to find $P[g(X) \leq y]$ since the required integral may be intractable.

Example 5: Suppose

$$g(x) = \begin{cases} x & -1 < x \leq 1 \\ -1 & x \leq -1 \\ +1 & x > 1 \end{cases} \tag{8}$$

and the PDF of X is $f_X(x) = \frac{1}{2}e^{-|x|}$, $x \in \mathbb{R}$. To find the CDF of Y, we need to first recognize that $-1 \le Y \le 1$. In fact, Y is a "clipped" version of X, with values of X outside the range $[-1, +1]$ being quantized to -1 or $+1$. Thus, for $y < -1$, $F_Y(y) = P[Y \le y] = 0$. Whenever $X \le -1$, we get $Y = -1$, therefore

$$\begin{aligned} P[Y = -1] &= P[X \le -1] = \int_{-\infty}^{-1} \frac{1}{2} e^x dx && (9) \\ &= \frac{1}{2} e^{-1}, && (10) \end{aligned}$$

which means there is a step discontinuity of magnitude $0.5e^{-1}$ at $y = -1$ in $F_Y(y)$. Next, for $-1 < y \le 0$, we have the equivalence $\{Y \le y\} = \{X \le y\}$, hence

$$\begin{aligned} F_Y(y) &= P[X \le y] = \int_{-\infty}^{y} \frac{1}{2} e^x dx && (11) \\ &= \frac{1}{2} e^y. && (12) \end{aligned}$$

For $0 < y < 1$, the equivalence $\{Y \le y\} = \{X \le y\}$ still holds, but this time we have to be careful about handling the magnitude sign, by splitting the integral of the PDF into two parts:

$$\begin{aligned} F_Y(y) &= \int_{-\infty}^{0} \frac{1}{2} e^x dx + \int_{0}^{y} \frac{1}{2} e^{-x} dx && (13) \\ &= 0.5 + 0.5 - 0.5e^{-y} = 1 - 0.5e^{-y}. && (14) \end{aligned}$$

Finally, $P[Y = 1] = P[X > 1] = 0.5e^{-1}$, resulting in a jump of that magnitude at $y = 1$. Therefore, the entire CDF is

$$F_Y(y) = \begin{cases} 0 & y < -1 \\ 0.5e^y & -1 \le y \le 0 \\ 1 - 0.5e^{-y} & 0 < y < 1 \\ 1 & y \ge 1 \end{cases} \tag{15}$$

The PDF is obtained by differentiating the CDF:

$$f_Y(y) = 0.5e^{-1}[\delta(y+1) + \delta(y-1)] + 0.5e^{-|y|}[u(y+1) - u(y-1)]. \tag{16}$$

The above unwieldy expression is much clearer when visualized on a plot. ■

1.4.1 Special Case 1: $Y = aX + b$

A commonly encountered function is $ax + b$. In this case, we have

$$\begin{aligned} F_Y(y) &= P[aX + b \le y] = \begin{cases} P[X \ge \frac{y-b}{a}] & a < 0 \\ P[X \le \frac{y-b}{a}] & a > 0 \end{cases} && (17) \\ &= \begin{cases} 1 - F_X\left(\frac{y-b}{a}\right) & a < 0 \\ F_X\left(\frac{y-b}{a}\right) & a > 0 \end{cases} && (18) \end{aligned}$$

(Strictly speaking, for the $a < 0$ case, the expression should be $1 - F_X((y-b)/a) + P[X = (y-b)/a]$, in case X is a mixed-type random variable.)

By differentiating the CDF, we obtain the PDF as

$$f_Y(y) = \begin{cases} -\frac{1}{a} f_X\left(\frac{y-b}{a}\right), & a < 0 \\ \frac{1}{a} f_X\left(\frac{y-b}{a}\right), & a > 0 \end{cases} = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right). \tag{19}$$

The final result comes from recognizing that $|a| = -a$ when a is negative, and $|a| = a$ when a is positive.

Example 6: Let $X \sim \mathcal{N}(\mu, \sigma^2)$, and $Y = aX + b$. Then from (19),

$$\begin{aligned} f_Y(y) &= \frac{1}{|a|\sqrt{2\pi\sigma^2}} \exp\left(-\frac{[(y-b)/a - \mu]^2}{2\sigma^2}\right) && (20) \\ &= \frac{1}{\sqrt{2\pi(a\sigma)^2}} \exp\left(-\frac{(y-b-a\mu)^2}{2(a\sigma)^2}\right). && (21) \end{aligned}$$

This is the PDF of an $\mathcal{N}(b + a\mu, (a\sigma)^2)$ random variable. Hence we have just shown that if X is Gaussian, then $aX + b$ is also Gaussian. ■

1.4.2 Special Case 2: $Y = X^2$

This is another commonly encountered function e.g. the power of a signal. The CDF of Y is found after noting the equivalence $\{Y \le y\} = \{-\sqrt{y} \le X \le \sqrt{y}\}$, when $y > 0$, which leads to

$$\begin{aligned} F_Y(y) &= P[-\sqrt{y} \le X \le \sqrt{y}] && (22) \\ &= F_X(\sqrt{y}) - F_X(-\sqrt{y}), \quad y > 0, && (23) \end{aligned}$$

assuming that X is continuous so that $P[X \ge -\sqrt{y}] = P[X > -\sqrt{y}]$. (If X is not continuous, we will have to add $P[X = -\sqrt{y}]$ into the above expression.) Therefore, the PDF of Y is

$$f_Y(y) = \frac{1}{2\sqrt{y}}[f_X(\sqrt{y}) + f_X(-\sqrt{y})], \quad y > 0. \tag{24}$$

Example 7: Let $X \sim \mathcal{N}(0, 1)$, and $Y = X^2$. Then (24) yields

$$\begin{aligned} f_Y(y) &= \frac{1}{\sqrt{y}} \cdot \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{y}{2}\right] && (25) \\ &= \frac{1}{\sqrt{2\pi y}} \exp\left[-\frac{y}{2}\right], \quad y > 0. && (26) \end{aligned}$$

1.5 Finding the PDF of Y from the PDF of X

The PDF of Y can be obtained directly from the PDF of X if the following two conditions are both met:

- The function $g(x)$ is such that $g'(x) = 0$ only at a countable number of points; and
- The random variable X is continuous.

If one or both of these conditions are violated, then we must resort to finding the CDF of Y. In the rest of this section, we assume that the two conditions are met.

As drawn in class, for a value of $y \in S_Y$, the equation $g(x) = y$ must have at least one root[2], i.e.

$$g(x_i) = y, \quad i = 1, 2, \ldots, n$$

where $n \geq 1$. For a positive but very small dy, the event $\{y < Y \leq y + dy\}$ is then equivalent to

$$\bigcup_{i=1}^{n} \{X \text{ in } |dx_i|\text{-neighborhood of } x_i\},$$

where $\frac{dy}{dx_i} = g'(x_i)$ is the gradient of $g(x)$ at $x = x_i$. (This gradient may be positive or negative, hence the necessity to express the equivalent event in X as a union of $|dx_i|$-neighborhoods.)

Given that dy is small, and that $g(x)$ satisfies the first condition, then $|dx_i|$ must be small. Therefore,

$$P[X \text{ in } |dx_i|\text{-neighborhood of } x_i] = f_X(x_i)|dx_i|, \tag{27}$$

since X satisfies the second condition, namely that it is continuous. At the same time, $P[y < Y \leq y+dy] = f_Y(y)dy$. Equating the probabilities of the two equivalent events (in Y and in X), we get

$$\begin{aligned} f_Y(y)dy &= \sum_{i=1}^{n} f_X(x_i)|dx_i| & (28)\\ \Rightarrow \quad f_Y(y) &= \sum_{i=1}^{n} f_X(x_i) \left|\frac{dx_i}{dy}\right| & (29)\\ &= \sum_{i=1}^{n} \frac{f_X(x_i)}{|g'(x_i)|}. & (30) \end{aligned}$$

Note that x_i can and should always be written in terms of y, so that the RHS appears in the final result as a function of y.

Example 8: Define $Y = X^2$, where X is a continuous r.v. with PDF $f_X(x)$. Then for any $y > 0$, $x^2 = y \Rightarrow x = \pm\sqrt{y}$, or in our notation above, $x_1 = -\sqrt{y}$, $x_2 = \sqrt{y}$.

[2] Otherwise, y cannot be an allowable value of $g(X)$, right?

The derivative of $g(x) = x^2$ is $g'(x) = 2x$, or $g'(x_1) = -2\sqrt{y} = -g'(x_2)$. Using (30), we get

$$f_Y(y) = \frac{f_X(-\sqrt{y})}{2\sqrt{y}} + \frac{f_X(\sqrt{y})}{2\sqrt{y}} \tag{31}$$

$$= \frac{1}{2\sqrt{y}}[f_X(-\sqrt{y}) + f_X(\sqrt{y})]. \tag{32}$$

This result coincides with (24). ■

Example 9: X has the PDF $f_X(x) = \lambda e^{-\lambda x}$, $x > 0$, and $Y = \sqrt{X}$. Find the PDF of Y.

Now we have $g(x) = \sqrt{x}$, and hence for any $y > 0$, $g(x) = y \;\Rightarrow\; x = y^2$ only, and $g'(y^2) = \frac{1}{2y}$. Using (30) we obtain

$$f_Y(y) = \frac{\lambda e^{-\lambda y^2}}{1/2y} \tag{33}$$

$$= 2\lambda y e^{-\lambda y^2}, \quad y > 0. \tag{34}$$

Y has a so-called Rayleigh distribution. ■

2 Two-Dimensional Random Vectors

2.1 Concept

In the previous section, we discussed $Y = g(X)$, where Y is a random variable defined from another r.v. X through a function $g(\cdot)$. This is a special case of what we will consider in the rest of the course – the case of two random variables X and Y, where Y is generally not a function of X.

As in our discussion of a single random variable X, we start with the abstract concept of a random experiment that produces outcomes ζ that belong in the sample space S. Instead of defining just one function $X(\zeta)$, we now define another one $Y(\zeta)$, where $Y(\zeta)$ is not necessarily generated from $X(\zeta)$ i.e. $Y(\zeta) \neq g(X(\zeta))$ for any $g(\cdot)$. So now we have a *random vector* (X, Y), rather than a scalar random variable.

We are interested in the *joint behaviour* of X and Y, most importantly whether and how X influences Y. A few examples will illustrate the problem.

2.2 Examples

Example A: Throw a dart at a round dart board with radius 40cm. Let X be the distance of the dart from the centre of the board, and let Y be the angle between the line from the centre of the board to the dart, and the x-axis. Then we have $X \in [0, 40)$, and $Y \in [0, 2\pi)$. Both X and Y are continuous random variables. Knowledge of X does not give us any information about Y e.g. even if we were told that the dart landed 10 cm away from the origin, we would be none the wiser about the angle it made with the x-axis. This is an example of a pair of independent random variables, which we will define properly later.

The range of possible values of the random vector (X, Y) is a rectangle within $\mathbb{R}^2$:

$$S_{X,Y} = \{(x, y) : 0 \le x < 40, 0 \le y < 2\pi\}.$$

Example B: Roll two dice. Let X be the larger of the two values seen, and let Y be the smaller of the two values. Then X and Y are both discrete random variables, with sample spaces $S_X = S_Y = \{1, 2, 3, 4, 5, 6\}$. The sample space of the random vector (X, Y) is not merely the Cartesian product of S_X and S_Y, because $X(\zeta) \ge Y(\zeta)$ for every outcome ζ. Instead,

$$S_{X,Y} = \{(x, y) : x \ge y, x \in S_X, y \in S_Y\}.$$

Unlike Example A, X and Y now contain information about each other e.g. if we know that $\{X = 2\}$ occurred, then Y can only be 1 or 2, i.e. $P[Y = k|X = 2] = 0 \ne P[Y = k]$ when $k > 2$. This type of conditional probability will be the basis for our definition of independence between random variables. ■

Example C: Two cities A and B are connected by four highways, numbered 401, 402, 403 and 404. The highway selected is defined as X, and the travel time between A and B is defined as Y. Now X is a discrete r.v., while Y is continuous, with $S_X = \{401, 402, 403, 404\}$ and $S_Y = (0, \infty)$. Assuming (reasonably) that the travel time Y is related to the highway selected (X), then we have X and Y being dependent. We would typically describe their relationship through a conditional distribution, such as $f_{Y|X}(y|x)$, which is the PDF of Y conditioned on $\{X = x\}$. We will discuss this idea in detail later. ■

2.3 Events

We have seen that events involving only X are subsets of $\mathbb{R}$ such as $\{X \le 2\}$, $\{-1 < X < 2\}$, etc. Events of interest related to the vector (X, Y) will naturally be subsets of $\mathbb{R}^2$, the two-dimensional Cartesian plane. For the three examples above, we can imagine the following events:

Example A – $\{X \cos Y > 10\}$, $\{X < 5\}$, $\{0 < Y \le \pi/2\}$.

Example B – $\{Y \le 2X\}$, $\{X + Y > 5\}$, $\{X - Y \le 2\}$.

Example C – $\{X = 401, Y \le 100\}$, $\{X \in \{403, 404\}, Y > 80\}$, $\{X = 402, 80 < Y < 100\}$.

Each of the above events has a physical interpretation. Can you work out their physical meaning?

3 Diagnostic Questions

1. Let $g(x) = \text{sgn}(x)$, and $Y = g(X)$. Suppose Y is uniform in $\{-1, +1\}$, what can we say about the distribution of X?

2. If $X \sim \mathcal{N}(0, 1)$, find the mean and variance of $Y = \sigma X + \mu$. From Example 6, Y is Gaussian – how would you generate 100 samples of Y if you have a way to generate 100 samples of X on a computer?

3. Find $P[X = 2, Y = 2]$ and $P[X = 2, Y = 3]$ in Example A above.