Indian Institute of Technology Bombay
Department of Electrical Engineering

Handout 17 EE 325 Probability and Random Processes
Generating Functions Sep 19,2013

1 Sum of Two Random Variables

We now consider the sum of two independent random variables, say X_1 and X_2, taking values in E_1 and E_2 respectively. We assume throughout that the sum is well defined and mostly limit ourselves to non-negative integer-valued random variables. However, many statements below can be appropriately generalized to real-valued discrete random variables.

First of all, $Y = X_1 + X_2$ is indeed a random variable, which can be checked from the basic definition of measurability from $(\Omega, \mathcal{F})$ to $(E, \mathcal{P}(E))$, where $\mathcal{F} = \mathcal{F}_1 \times \mathcal{F}_2$, $\mathcal{F}_i$ is the sigma-field with respect to which X_i is measurable. We also know from the linearity of expectation that

$$\mathbb{E}[Y] = \mathbb{E}[X_1] + \mathbb{E}[X_2].$$

Similarly

$$\sigma_Y^2 = \sigma_{X_1}^2 + \sigma_{X_2}^2.$$

However, we know that the random variable Y is specified in terms of its probability distribution function $P(Y = y), \forall y \in E$. It turns out that we can find $P(Y = y)$ in terms of $P(X_i = x_i)$ in a straightforward fashion.

$$\begin{aligned} P(Y = y) &= \sum_{x_1} P(X_1 = x_1, X_2 = y - x_1) \\ &= \sum_{x_1} P(X_1 = x_1) P(X_2 = y - x_1) \end{aligned}$$

The last equation resembles the traditional *convolution* operation in signals and systems. Recall that $g(x) = f(x) * h(x)$ implies

$$g(x) = \int f(u)g(x-u)du = \int f(x-u)g(u)du, \qquad (1)$$

where the integral is replaced by a summation in the discrete case. We highlight this result for future use.

> The probability distribution of the sum of two independent discrete random variables is the convolution of the individual distributions. We will denote this as $P_Y = P_{X_1} * P_{X_2}$.

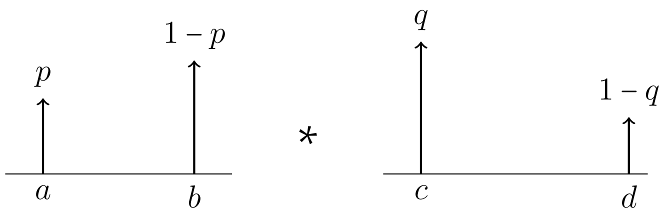
Later we will see that the above formula holds true for the sum of real valued random variables too.

Example 1 *Let X_1 be a RV such that $P(X_1 = a) = p$ and $P(X_1 = b) = 1 - p$. Consider an independent random variable X_2 with $P(X_2 = c) = q$ and $P(X_2 = d) = 1 - q$. Find the probability distribution of $Y = X_1 + X_2$ and sketch it.*

Solution: For the ease of illustration, assume that a, b, c, d are distinct numbers and $a + d \neq b + c$. Evidently, the possible values of Y are in $\{a+c, a+d, b+c, b+d\}$. Furthermore,

$$
\begin{aligned}
P(Y = a + c) &= P(X_1 = a)P(X_2 = c) = pq \\
P(Y = a + d) &= P(X_1 = a)P(X_2 = d) = p(1 - q) \\
P(Y = b + c) &= P(X_1 = b)P(X_2 = c) = (1 - p)q \\
P(Y = a + c) &= P(X_1 = b)P(X_2 = d) = (1 - p)(1 - q)
\end{aligned}
$$

So the convolution formula in (1) as such was not really needed, nevertheless let us illustrate pictorially that convolution will indeed give this result.



Distribution of X_1 Distribution of X_2

Notice that convolution of any function $f(x)$ and an impulse of magnitude α at position t_0 will result is $\alpha f(x - t_0)$, i.e. the same function scaled by the impulse magnitude and shifted to the position of the impulse. By linearity, the convolution with two impulses can be written as the convolution on individual impulses and then adding the results. The resulting convolution of our example is illustrated in Figure 1.

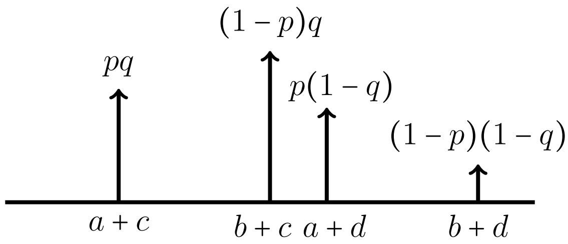


Figure 1: Distribution $P(Y = y)$

■

2 Generating Functions

Writing a result as a convolution has many advantages. Foremost of these is the conjugal relationship of convolution with multiplication in the transform domain, popularly known as convolution-multiplication theorem. For example, Laplace Transform, Fourier Transform etc for continuous-time functions, and the so called $Z-$Transform for discrete-time signals. We do not need the deeper aspects of these theories, this you will learn in coming semesters, but some superficial properties are enough for our purpose. This, admittedly, is a little extra effort, but certainly very beneficial. We will focus on the $Z-$Transform, which is defined for a sequence $\alpha_n, n \in \mathbb{N} \bigcup \{0\}$ as,

$$g(z) = \sum_k \alpha_k z^k.$$

where z takes values in the complex plane. Notice that the RHS is nothing but the power series expansion of $g(z)$. All we are doing is to find the function $g(z)$ with the given 'polynomial' coefficients, albeit of possibly unbounded degree. For those who are familiar with digital filters, the unbounded degree case corresponds to what is known as IIR filters (infinite impulse response). Since z takes complex values (remember the equivalent e^s term in Laplace transform), we have to define expectation of a complex random variable. We use the natural extension

$$\mathbb{E}[X_R + jX_I] = \mathbb{E}[X_R] + j\mathbb{E}[X_I],$$

where X_R and X_I are the real and imaginary parts respectively of the given complex variable.

Definition 1 *The generating function $g_X(z)$ of a non-negative integer valued random variable X is defined as*

$$g_X(z) = \mathbb{E}[z^X] = \sum_{k\geq 0} P(X = k)z^k.$$

The generating function, also denoted as GF, completely specifies the probability distribution. This is clear by noticing that once we expand $g_X(z)$ as power series, the coefficient of the k^{th} term is indeed $P(X = k)$. So if $g_X(z)$ is all about obtaining $P(X = k)$, why take the extra trouble to define it? It turns out that $g_X(z)$ is computationally more useful than the distribution function when it comes to sums of independent random variables.

Theorem 1 *Consider independent random variables X and Y. Then*

$$g_{X+Y}(z) = g_X(z)g_Y(z)$$

Solution: Notice that this is re-stating the convolution-multiplication theorem.

$$\begin{aligned} g_{X+Y}(z) &= \mathbb{E}z^{X+Y} \\ &= \mathbb{E}z^X z^Y \\ &= \mathbb{E}z^X \mathbb{E}z^Y \\ &= g_X(z)g_Y(z), \end{aligned}$$

where the third inequality used the fact that X and Y are independent. Keep in mind that the statement in general may not be true without assuming independence.

■

Thus it is simple to compute the generating function of independent sums, and from this we can easily obtain the distribution of the sum. We claim without proving that we can *invert* the generating function to obtain the distribution. While the matter of exact inversion can be a bit subtle in general (as in the case of Inverse Fourier Transform), let us not worry about this, pathological cases are seldom encountered. We are not going to deal with elaborate inversion formulas or mechanisms, but use our knowledge on a case by case basis, i.e. we know the generating functions of several widely used discrete random variables, and we will simply identify the distribution of X by observing $g_X(z)$.

Example 2 *Find the GF of a Binomial(n,p) random variable.*

Solution:

$$\begin{aligned} g(z) &= \sum_{k=0}^{n} P(X = k)z^k \\ &= \sum_{k=0}^{n} \binom{n}{k} (pz)^k (1-p)^{n-k} \\ &= (1 - p + pz)^n \end{aligned}$$

Example 3 *Find the GF of a Poisson random variable of parameter λ.*

We know that for a $Poisson(\lambda)$

$$P(X = k) = e^{-\lambda k}\frac{\lambda^k}{k!}, k \geq 0$$

$$\begin{aligned} g(z) &= \sum e^{-\lambda}\frac{(\lambda z)^k}{k!} \\ &= e^{\lambda(z-1)} \end{aligned}$$

Example 4 *What is the distribution of $X_1 + X_2$ if X_1 and X_2 are independent, and $X_1 \sim Binomial(n_1, p)$, $X_2 \sim Binomial(n_2, p)$*

Solution: The easiest way is to find the GF of $X_1 + X_2$.

$$\begin{aligned} g_{X_1+X_2}(z) &= g_{X_1}(z)g_{X_2}(z) \\ &= (1-p+pz)^{n_1}(1-p+pz)^{n_2} \\ &= (1-p+pz)^{n_1+n_2} \end{aligned}$$

Thus the GF of $X_1 + X_2$ corresponds to $Binomial(n_1 + n_2, p)$. Observe that this does not hold if the second parameter p was not identical. Equivalently, identical coins were used in the generation of X_1 and X_2 (independently).

■

Example 5 *What is the distribution of $X_1 + X_2$ if X_1 and X_2 are independent, and $X_1 \sim Poisson(\lambda_1)$, and $X_2 \sim Poisson(\lambda_2)$.*

Solution: Unlike the previous question, here the random variables can have totally different parameters. Proceeding as above,

$$\begin{aligned} g_{X_1+X_2}(z) &= g_{X_1}(z)g_{X_2}(z) \\ &= e^{\lambda_1(z-1)}e^{\lambda_2(z-1)} \\ &= e^{(\lambda_1+\lambda_2)(z-1)}, \end{aligned}$$

which corresponds to a Poisson process of parameter $\lambda_1 + \lambda_2$.

■

Repeating the above argument, we have the following theorem.

Theorem 2 *The sum of k independent Poisson RVs of respective parameters $\lambda_i, 1 \leq i \leq k$ is Poisson distributed with parameter $\sum_{i=1}^k \lambda_i$.*

Another advantage of having the GF is that we can evaluate quantities like the mean and variance directly, without recourse to finding the probability distribution.

Theorem 3 *Given a GF $g_X(z)$ which is twice differentiable,*

$$\mathbb{E}[X] = g'(1)$$
$$\mathbb{E}[X^2] = g''(1) + g'(1)$$

Solution: The differentiability assumption is technical, and there are techniques to avoid it. The proof is simple.

■

3 Conditional Expectation

We have already learned about expectations with respect to a distribution function. For a function of two random variables,

$$\mathbb{E}[g(X,Y)] = \sum_{x,y} g(x,y)P(X = x, Y = y).$$

It is possible to extend this definition to compute expectation with respect to conditional probabilities. Recall that for every $Y = y$, such that $P(Y = y) > 0$, $P(X|Y = y)$ is a probability distribution function. In particular,

$$\sum_{x \in E_1} P(X|Y = y) = 1.$$

Definition 2 *Let X and Y takes values in E_1 and E_2 respectively. Let the function $g : E_1 \times E_2 \to \mathbb{R}$, be either non-negative or $\mathbb{E}|g(X,Y)| \leq \infty$. Then for all $y \in E_2$ such that $P(Y = y) > 0$, the conditional expectation given $Y = y$, is defined as*

$$\mathbb{E}[g(X,Y)|Y = y] = \sum_{x \in E_1} g(x,y)P(X = x|Y = y)$$

The conditional expectation $\mathbb{E}[g(X,Y)|Y = y]$ is a function of y, let us call it $\psi(y)$. Considering Y as a random variable, $\psi(Y)$ is a random variable as it is function of Y.

Definition 3 *The conditional expectation defined as*

$$\psi(Y) = \mathbb{E}[g(X,Y)|Y],$$

is a random variable.

In order to avoid the confusion in cases where $P(Y = y) = 0$, let us make the convention that zero multiplied by *anything* is zero. i.e. if we observe a zero in a product, let us not care whether the rest of the product is well-defined or not.

The properties of expectation like linearity and monotonicity hold for conditional expectation also.

Example 6 *Let X_1 and X_2 be identical binomial random variables of size N and parameter p. Find the conditional expectation $\mathbb{E}[X_1|X_1 + X_2]$.*

Solution: Notice the form $\mathbb{E}[X_1|Y]$ where $Y = f(X_1, X_2)$. Thus, what we obtain as conditional expectation will be a random variable in terms of Y.

$$\begin{aligned}
\mathbb{E}[X_1|X_1 + X_2 = n] &= \sum_{i=0}^{n} iP(X_1 = i|X_1 + X_2 = n) \\
&= \sum_{i=0}^{n} i\frac{P(X_1 = i)P(X_2 = n - i)}{P}(X_1 + X_2 = n) \\
&= \sum_{i=0}^{n} i\frac{\binom{n}{i}\binom{n}{n-i}}{\binom{2N}{n}} \\
&= \frac{n}{2}
\end{aligned}$$

Since this is true for any n that Y can take, we deduce that

$$\mathbb{E}[X_1|X_1+X_2] = \frac{X_1+X_2}{2}.$$

■

We have defined $\psi(Y) = \mathbb{E}[g(X,Y)|Y]$ to emphasize that it is a random variable. Now we can think of taking its expectation with respect to the distribution of Y. When there are multiple expectations involved, it is instructive to add a subscript to the expectation, denoting which variable is being integrated. For example $\psi(Y) = \mathbb{E}_x[g(X,Y)|Y]$, saying that the summation is over $x \in E_1$. We now show a very important result.

Theorem 4

$$\mathbb{E}_y\mathbb{E}_x[g(X,Y)|Y=y] = \mathbb{E}_{x,y}g(X,Y)$$

Proof: Using the expression $\mathbb{E}\psi(Y) = \sum \psi(y)P(Y=y)$,

$$\begin{aligned}
\mathbb{E}_y\mathbb{E}_x[g(X,Y)|Y=y] &= \sum_{y\in E_2} P(Y=y)\mathbb{E}[g(X,Y)|Y=y] \\
&= \sum_{y\in E_2} P(Y=y) \sum_{x\in E_1} g(x,y)P(X=x|Y=y) \\
&= \sum_{x,y} P(X=x,Y=y)g(x,y) \\
&= \mathbb{E}g(X,Y).
\end{aligned}$$

This result is very useful, and leads to a famous identity called Wald's identity.

4 Random Sums and Wald's Identity

Consider n random variables $X_1, \cdots, X_n$. We know how to calculate the moments of the sum $Y = \sum_{i=1}^n X_i$. However, there are cases where the number of random variables to be summed itself is chosen in a random fashion. Consider

$$Y = \sum_{i=1}^{T} X_i,$$

where T is a random number independent of $X_1, \cdots, X_n$. In order to compute the mean of Y, a famous formula known as Wald's identity comes to our rescue.

Theorem 5 *Consider random variables X_i, each having the same mean $\mathbb{E}[X]$.*

$$\mathbb{E}[Y] = \mathbb{E}[X]\mathbb{E}[T].$$

Proof: From the previous section,

$$\mathbb{E}[Y] = \mathbb{E}_t\mathbb{E}_y[Y|T].$$

Thus,

$$\begin{aligned}
\mathbb{E}[Y] &= \mathbb{E}_t T\,\mathbb{E}[X] \\
&= \mathbb{E}[T]\mathbb{E}[X].
\end{aligned}$$