RANDOM VARIABLES

Mauro Díaz

4.3

Exercise 4.3 (Convolutions) Convolutions are probability distributions that correspond to sums of independent random variables.

a. Let X and Y be independent discrete variables. Find the PMF of $Z = X + Y$. Hint: Use the law of total probability.

b. Let X and Y be independent continuous variables. Find the PDF of $Z = X + Y$. Hint: Start with the CDF.

(a) For the variables X, Y : $\exists\, B_X, B_Y \in \mathcal{B}(\mathbb{R})$ countable s.t. $P_X(B_X) = P_Y(B_Y) = 1$ and as countable, there is a bijection with $\mathbb{N}$ for every point on both subsets, and we can write $B_X = \bigcup_{i \in \mathbb{N}} \{x_i\}$, $B_Y = \bigcup_{i\in\mathbb{N}} \{y_i\}$ where $x_i \in B_X$, $y_i \in B_Y$.

Now, we can use the law of total probability for $Z = X + Y$:

$$P(Z=z) = P(X+Y=z) = \sum_{i=1}^{\infty} P(X+y_i = z \mid Y = y_i)\, P(Y=y_i) =$$

X and Y independent

$$= \sum_{i=0}^{\infty} P(X + y_i = z)\, P(Y=y_i) = \sum_{i=0}^{\infty} P(X = (z - y_i))\, P(Y = y_i)$$

(b) Lets define f, g the PDFs of X and Y respectively:

total probability law

independents

$$F_z(y) = P(Z<z) = P(X+Y<z) = \int_{\mathbb{R}} P(X+y<z \mid Y=y)\, P(Y=y)\, dy =$$

by working in dy : $P(y \le Y \le y+dy) = g(y)$

$$= \int_{\mathbb{R}} P(X+y<z)\, P(Y=y)\, dy = \int_{\mathbb{R}} P(X<z-y)\, P(Y=y)\, dy =$$

$$= \int_{\mathbb{R}} \left(\int_{-\infty}^{z-y} f(x)\, dx \right) g(y)\, dy$$

Now that we have an expression for F_z, we can derivate to get the PDF:

We will define $h(z,y) = \int_{-\infty}^{z-y} f(x)\, dx$ and by the fundamental theorem of calculus we have $\frac{\partial}{\partial z} h(z,y) = f(z-y) \cdot \frac{\partial}{\partial z}(z-y) = f(z-y)$.

Now, applying Leibniz's rule:

$$\frac{\partial}{\partial z} F_z(z) = \frac{\partial}{\partial z} \int_{\mathbb{R}} \underbrace{\int_{-\infty}^{z-y} f(x)}_{h(z,y)}\, g(y)\, dy = \int_{\mathbb{R}} \left(\frac{\partial}{\partial z} h(z,y) \right) g(y)\, dy =$$

$$= \boxed{\int_{\mathbb{R}} f(z-y)\, g(y)\, dy}$$

4.7.

Exercise 4.7 (Negative binomial random variable) A variable with PMF

$$p(k) = \binom{k+r-1}{k}(1-p)^r p^k$$

is a negative binomial random variable with support in non-negative integers. It has two parameters $r > 0$ and $p \in (0, 1)$. We denote

$$X | r, p \sim \mathrm{NB}(r, p).$$

a. Let us reparameterize the negative binomial distribution with $q = 1 - p$. Find the PMF of $X \sim \mathrm{NB}(1, q)$. Do you recognize this distribution?

b. Show that the sum of two negative binomial random variables with the same p is also a negative binomial random variable. Hint: Use the fact that the number of ways to place n indistinct balls into k boxes is $\binom{n+k-1}{n}$.

(a)

Lets do the substitution in $p(k)$ with $r = 1$, $q = 1-p \Rightarrow p = 1-q$:

$$p(k) = \binom{k+r-1}{k}(1-p)^r p^k = \binom{k}{k} q (1-q)^k = q(1-q)^k$$ --> Geometric distribution with probab. of success q.

(b) We will use the total probability law as in the exercise 4.3. Hence, we have: $X \sim NB(r_1, p)$, $Y \sim NB(r_2, p)$ and $Z = X + Y$:

4.3. and $k<0$ no sense

$$P(Z = z) = \sum_{k=0}^{\infty} P(X = z-k) P(Y = k) =$$

no sense $k > z$

$$= \sum_{k=0}^{z} P(X = z-k) P(Y=k) =$$

$$= \sum_{k=0}^{z} \binom{z-k+r_1-1}{z-k}\binom{k+r_2-1}{k}(1-p)^{r_1+r_2} p^z =$$

$$= (1-p)^{r_1+r_2} p^z \sum_{k=0}^{z} \binom{z-k+r_1-1}{z-k}\binom{k+r_2-1}{k} =$$

$z-k$ balls into r_1 boxes

k balls into r_2 boxes

$\forall k \in \{0, \dots, z\}$ we partition the number z of balls in "2 parts" (k and $(z-k)$) and we measure how to put them in r_2 and r_1 boxes

Summing over all the type of partitions, we are counting all the ways to put z balls in r_1 boxes and z balls in r_2 boxes

$$= (1-p)^{r_1+r_2} p^z \binom{z+r_1+r_2-1}{z} \sim NB(r_1+r_2, p) \;!!$$

Remark: It is (for me) easier to see for $\binom{n+k-1}{n}$ that we have $n+k-1$ position for putting the n balls and $k-1$ barriers for create the boxes.

4.14.

Exercise 4.14 (Cantor distribution) The Cantor set is a subset of $[0,1]$, which we create by iteratively deleting the middle third of the interval. For example, in the first iteration, we get the sets $[0,\frac{1}{3}]$ and $[\frac{2}{3},1]$. In the second iteration, we get $[0,\frac{1}{9}]$, $[\frac{2}{9},\frac{1}{3}]$, $[\frac{2}{3},\frac{7}{9}]$, and $[\frac{8}{9},1]$. On the n-th iteration, we have

$$C_n = \frac{C_{n-1}}{3} \cup \left(\frac{2}{3} + \frac{C_{n-1}}{3}\right),$$

where $C_0 = [0,1]$. The Cantor set is then defined as the intersection of these sets

$$C = \cap_{n=1}^{\infty} C_n.$$

It has the same cardinality as $[0,1]$. Another way to define the Cantor set is the set of all numbers on $[0,1]$, that do not have a 1 in the ternary representation $x = \sum_{n=1}^{\infty} \frac{x_i}{3^i}, x_i \in \{0,1,2\}$.

A random variable follows the Cantor distribution, if its CDF is the Cantor function (below). You can find the implementations of random number generator, CDF, and quantile functions for the Cantor distributions at https://github.com/Henrygb/CantorDist.R.

a. Show that the Lebesgue measure of the Cantor set is 0.

b. (**Jagannathan**) Let us look at an infinite sequence of independent fair-coin tosses. If the outcome is heads, let $x_i = 2$ and $x_i = 0$, when tails. Then use these to create $x = \sum_{n=1}^{\infty} \frac{x_i}{3^i}$. This is a random variable with the Cantor distribution. Show that X has a singular distribution.

(a) We will analize first of all, the measure of C_n $\forall n \in \mathbb{N}$

We start with $C_0 = [0,1]$, so C_1 will be C_0 removing the middle third $(1/3, 2/3)$, finally having $C_1 = [0, 1/3] \cup [2/3, 1] \Rightarrow 2$ intervals of length $1/3$.

- C_2 : every interval of C_1 removing the middle third $\Rightarrow$ we will have 4 intervals of length $1/9$ $\Rightarrow \lambda(C_2) = \frac{4}{9} = \left(\frac{2}{3}\right)^2$

$\vdots \qquad \vdots$

- C_n : will have 2^n intervals of length $1/3^n$ $\Rightarrow \lambda(C_n) = \frac{2^n}{3^n} = \left(\frac{2}{3}\right)^n$

Using now that $C_{n+1} \subseteq C_n$ and that $C = \bigcap_{n=0}^{\infty} C_n$ we can use the limits:

$$\lambda(C) = \lambda\left(\bigcap_{n=0}^{\infty} C_n\right) = \lim_{n\to\infty} \lambda(C_n) = \lim_{n\to\infty} (2/3)^n = 0$$

(b) X will be, by the statements, a random variable that modelizes an infinite seq. of indep. fair-coin tosses. We will have the correspondence between Head (H) and Tail (T) with $x = \sum_{i=1}^{\infty} \frac{x_i}{3^i} \in C$ such that:

H in i-toss $\Leftrightarrow x_i = 2$, T in j-toss $\Leftrightarrow x_j = 0$.

First of all, this P_X is continuous respect to λ because for every singleton:

$$P_X(x) = P_X\left(\sum_{i=0}^{\infty} \frac{x_i}{3^i}\right) = \lim_{n\to\infty} P_X\left(\sum_{i=0}^{n} \frac{x_i}{3^i}\right) = \lim_{n\to\infty} P_X(\text{n-sequence of tosses}) = \lim_{n\to\infty} \left(\frac{1}{2}\right)^n = 0.$$

X is a singular distribution because:

$P_X(C) = 1$ and $\lambda(C) = 0$.

represent the set of all infinite seq. of fair-coin tosses (probability of 1, there is always 1 inf. seq.)

4.18.

Exercise 4.18 (Probability integral transform) This exercise is borrowed from Wasserman. Let X have a continuous, strictly increasing CDF F. Let $Y = F(X)$.

a. Find the density of Y. This is called the probability integral transform.

b. Let $U \sim \mathrm{Uniform}(0, 1)$ and let $X = F^{-1}(U)$. Show that $X \sim F$.

(a) We will first find the CDF of Y using the $Y = F(X)$ property.

F is continuous and strictly increasing $\Rightarrow$ F^{-1} exists and is also strictly increasing

$F_Y(y) = P(Y < y) = P(F_X(X) < y) = P(X < F_X^{-1}(y)) = F_X(F_X^{-1}(y)) = y$

Hence, if we have $F_Y(y) = y \Rightarrow f_Y(y) = \frac{\partial}{\partial y} F_Y(y) = 1$

(b) $U \sim U(0,1)$ so $F_U(y) = y$ and $f_U(y) = 1$.

We can now do almost the same procedure of (a):

$F_X(x) = P(X < x) = P(F^{-1}(U) < x) = P(U < F(x)) = F_U(F(x)) = F(x)$

same prop. of inverse

So X has CDF $F \Rightarrow X \sim F$.