Discrete Random Processes (EDRP)

Lecture 7

A basic probability refresher - part 3

Consider a *discrete* random variable X - that is one that takes values in a finite set, or in the set $\mathbb{N}$ of natural numbers, or in the set of integers $\{0, \pm 1, \pm 2, \ldots\}$. The distribution of X is described by the probability mass function (pmf)

$$x \mapsto \mathbb{P}(X = x).$$

If $A \subset \mathbb{R}$ and

$$S_X = \{x : \mathbb{P}(X = x) \neq 0\}$$

(this is the **support** of the distribution of X), then

$$\mathbb{P}(X \in A) = \sum_{x \in A \cap S} \mathbb{P}(X = x).$$

The **expectation**, or **mean**, of a discrete random variable X is defined as

$$\mathbb{E}X = \sum_{x \in S_X} x \cdot \mathbb{P}(X = x).$$

- The expectation is a *weighted average* of the values of X, with weights given by the probabilities.
- Intuitively, the expectation of X is the *long-run average value* of X over repeated trials.

If g is a function and X is a random variable, then $Y = g(X)$ is a **function of a random variable**, which itself is a random variable that takes the value $g(x)$ whenever X takes the value x. A useful formula for computing the expectation of a function of a random variable is

$$\mathbb{E}g(X) = \sum_{x \in S_X} g(x) \cdot \mathbb{P}(X = x).$$

Example 1

We roll a die and draw a square with the side of the length equal the number the die came up. What is the mean area of the square?

The **variance** of a random variable is a measure of variability or discrepancy from the mean. It is defined as

$$\operatorname{Var} X = \mathbb{E}\left[(X - \mathbb{E}X)^2\right] = \sum_{x \in S_X} (x - \mathbb{E}X)^2 \cdot \mathbb{P}(X = x).$$

A computationally useful formula is

$$\operatorname{Var} X = \mathbb{E}\left(X^2\right) - (\mathbb{E}X)^2 .$$

Example 2

Let $p \in (0,1)$ and $t \geq 0$. Suppose that $p = \mathbb{P}(X = \pm t)$ and $\mathbb{P}(X = 0) = 1 - 2p$. Compute $\mathbb{E}X$ and $\operatorname{Var} X$.

The **standard deviation** of a random variable is defined as

$$\mathrm{SD}(X) = \sqrt{\operatorname{Var} X}.$$

Probability generating functions

For a discrete random variable X taking values in $\{0, 1, \ldots\}$, the **probability generating function of X** *(pgf of X) is the function*

$$G_X(s) = \mathbb{E}\left(s^X\right) = \sum_{k=0}^{\infty} s^k \cdot \mathbb{P}(X = k).$$

So

$$G_X(s) = \mathbb{P}(X = 0) + s \cdot \mathbb{P}(X = 1) + s^2 \cdot \mathbb{P}(X = 2) + s^3 \cdot \mathbb{P}(X = 3) \ldots$$

– the pgf is a *power series* whose coefficients are probabilities.

Example 3

Compute the pgf of X if

$$\mathbb{P}(X = 0) = \mathbb{P}(X = 1) = \mathbb{P}(X = 2) = \frac{1}{3}.$$

However, pgf's are not always polynomials:

Example 4

Assume that X has a geometric distribution with parameter $1/6$. Find the pgf of X.

$G_X \equiv G_Y$ implies equality of distributions

The generating function represents the distribution of a discrete random variable as a power series.
If two power series are equal, then they have the same coefficients.
Hence, if two discrete random variables X and Y have the same pgf's, that is

$$G_X(s) = G_Y(s) \quad \forall s,$$

then

$$\mathbb{P}(X = k) = \mathbb{P}(Y = k) \quad \forall k \in \{0, 1, 2, \ldots\},$$

so X and Y have the same distribution.

Recovering probabilities from pgf

Probabilities for X can be obtained from the generating function by successive differentiation:

$$
\begin{aligned}
G_X(0) &= \mathbb{P}(X=0), \\
G'_X(0) &= \sum_{k=1}^{\infty} k s^{k-1} \mathbb{P}(X=k) \bigg|_{s=0} = \mathbb{P}(X=1), \\
G''_X(0) &= \sum_{k=2}^{\infty} k(k-1) s^{k-2} \mathbb{P}(X=k) \bigg|_{s=0} = 2\mathbb{P}(X=2),
\end{aligned}
$$

and so on.

Recovering probabilities from pgf – cont'd

In general,

$$G_X^{(j)}(0) = \sum_{k=j}^{\infty} k(k-1)\cdot \ldots \cdot (k-j+1)s^{k-j}\mathbb{P}(X=k)\Bigg|_{s=0}$$
$$= j!\mathbb{P}(X=j),$$

(where $G^{(j)}$ denotes the jth derivative of G), and thus

$$\mathbb{P}(X=j) = \frac{G_X^{(j)}(0)}{j!}, \quad j = 0, 1, 2, \ldots.$$

Recovering probabilities from pgf – cont'd

Example 5

A random variable X has pgf $G_X(s) = (1 + s + s^2)/3$. Find the distribution of X.

Pgfs and sums of independent random variables

Pgfs are useful tools for working with sums of independent random variables. Assume that $X_1, \ldots, X_n$ are independent, and let

$$U = X_1 + \ldots + X_n.$$

The pgf of U is

$$G_U(s) = \mathbb{E}\left(s^U\right) = \mathbb{E}\left(s^{X_1+\ldots+X_n}\right) = \mathbb{E}(s^{X_1} \cdot \ldots \cdot s^{X_n}) =$$
$$= \mathbb{E}\left(s^{X_1}\right) \cdot \ldots \cdot \mathbb{E}\left(s^{X_n}\right) = G_{X_1}(s) \cdot \ldots \cdot G_{X_n}(s),$$

where the fourth equality is by independence. Thus, the pgf of an independent sum is the product of the individual generating functions.

If the X_i's are also identically distributed, then

$$G_{X_1+\ldots+X_n}(s) = G_{X_1}(s) \cdot \ldots \cdot G_{X_n}(s) = \left[G_{X_1}(s)\right]^n.$$

Pgfs and sums of independent random variables – cont'd

Example 6

Suppose X_1 and X_2 are independent random variables uniformly distributed on $\{0, 1\}$. Find the distribution of $X_1 + X_2$. Compute the pgfs of X_1, X_2, and $X_1 + X_2$.

Pgfs and moments

The pgf of X can be used to find the expected value, variance, and higher moments of X. Observe that

$$G'_X(1) = \mathbb{E}\left(Xs^{X-1}\right)\Big|_{s=1} = \mathbb{E}(X).$$

Also,

$$G''_X(1) = \mathbb{E}\left(X(X-1)s^{X-2}\right)\Big|_{s=1} = \mathbb{E}(X(X-1)) = \mathbb{E}(X^2) - \mathbb{E}(X),$$

so

$$\operatorname{Var} X = \mathbb{E}(X^2) - \mathbb{E}^2(X) = G''_X(1) + G'_X(1) - \left(G'_X(1)\right)^2.$$

Example 7

A random variable X has pgf $G_X(s) = (1 + s + s^2)/3$. Compute $\mathbb{E}X$ and $\operatorname{Var} X$.

Branching processes

Branching processes are a class of stochastic processes that model the growth of populations.

Assume that we have a population of individuals, each of which independently produces a random number of children according to a probability distribution

$$\mathbf{a} = (a_0, a_1, \ldots).$$

That is, an individual gives birth to k children with probability a_k, for $k \geq 0$, *independently of other individuals*. We call $\mathbf{a}$ the **offspring distribution**.

The population changes from generation to generation. For $n \geq 0$ let

Z_n be the size (number of individuals) of the nth generation.

Assume $Z_0 = 1$ (that is, the population starts with one individual).

Definition 1

The sequence $(Z_n)_{n \in \mathbb{N}_0} = (Z_0, Z_1, Z_2, \ldots)$ is a **branching process**.

What does a realization of such a process look like?

A branching process $(Z_n)_{n\in\mathbb{N}_0}$ is a Markov chain since the size of a generation only depends on the size of the previous generation and the number of their offspring. If Z_n is given, then the size of the next generation Z_{n+1} is *independent* of $Z_0, \ldots, Z_{n-1}$.

Assumptions

We are going to assume that

$$0 < a_0 < 1.$$

Observe that

- if $a_0 = 0$, then the population only grows and 0 is not in the state space,
- if $a_0 = 1$, then $Z_n = 0$, for all $n \geq 1$.

By the way:

if $Z_n = 0$, we say that **the process has become extinct by generation** n.

We will also assume that there is a positive probability that an individual gives birth to more than one offspring, that is,

$$a_0 + a_1 < 1.$$

Mean and variance of generation size

In a branching process, the **size of the *n*th generation** is the sum of the total offspring of the individuals of the previous generation. That is,

$$Z_n = \sum_{i=1}^{Z_{n-1}} X_i,$$

where X_i denotes the number of children born to the ith person in the $(n-1)$th generation. Because of the independence assumption, $X_1, X_2, \ldots,$ is an i.i.d. sequence with common distribution $\mathbf{a}$. Furthermore, Z_{n-1} is independent of $X_1, X_2, \ldots$.

Thus Z_n is a *random sum* of i.i.d. random variables.

Mean and variance of generation size - cont'd

Let μ be the mean of the offspring distribution **a**, that is

$$\mu = \sum_{k=0}^{\infty} k a_k.$$

Let σ^2 denote the variance of the offspring distribution, that is

$$\sigma^2 = \left(\sum_{k=0}^{\infty} k^2 a_k \right) - \mu^2.$$

We will be interested in the mean $\mathbb{E} Z_n$ and variance $\operatorname{Var} Z_n$ of the size of the nth generation.

Mean and variance of generation size - cont'd

Theorem 2

The mean and variance of the size of the nth generation (with $n \geq 0$) are

$$\begin{aligned} \mathbb{E} Z_n &= \mu^n, \\ \operatorname{Var} Z_n &= \begin{cases} \sigma^2 n, & \text{if } \mu = 1, \\ \sigma^2 \mu^{n-1} \frac{\mu^n - 1}{\mu - 1}, & \text{if } \mu \neq 1. \end{cases} \end{aligned}$$

Mean generation size in the long term

We have then three cases for the long-term expected generation size:

$$\lim_{n\to\infty} \mathbb{E}Z_n = \lim_{n\to\infty} \mu^n = \begin{cases} 0, & \text{if } \mu < 1, \\ 1, & \text{if } \mu = 1, \\ +\infty, & \text{if } \mu > 1. \end{cases}$$

Definition 3

A branching process is said to be

- **subcritical** if $\mu < 1$,
- **critical** if $\mu = 1$,
- and **supercritical** if $\mu > 1$.

Mean generation size in the long term - cont'd

For a subcritical branching process, mean generation size declines exponentially to zero. For a supercritical process, mean generation size exhibits long-term exponential growth.

The limits suggest three possible regimes depending on μ:

- long-term extinction, or
- stability, or
- boundless growth.

However, behavior of the *mean* generation size does not tell the whole story.

Mean and variance of the generation size in the long term

To explore the process of extinction, let us additionally consider the variance

$$\mathrm{Var}\, Z_n = \begin{cases} \sigma^2 n, & \text{if } \mu = 1, \\ \sigma^2 \mu^{n-1} \frac{\mu^n - 1}{\mu - 1}, & \text{if } \mu \neq 1. \end{cases}$$

- In the subcritical case, both the mean and variance of generation size tend to 0.
- In the critical case, $\mathbb{E} Z_n = 1$ for all n, but the variance is a linearly growing function of n.
- In the supercritical case, the variance grows exponentially large. The large variance suggests that in some cases both extinction and boundless growth are possible outcomes.

In fact, it turns out that

- in the subcritical and critical cases ($\mu \leq 1$), the population becomes extinct with probability 1,
- in the supercritical case ($\mu > 1$), the probability that the population eventually dies out is less than one, but typically greater than zero.

Branching processes and probability generation functions

We will use probability generating functions to analyze those issues. Let

$$G(s) = \sum_{k=0}^{\infty} s^k a_k$$

be the generating function of the offspring distribution.

Theorem 4

The probability of eventual extinction of a branching process is the smallest positive root of the equation $s = G(s)$.

Clearly, $s = 1$ satisfies the equation, but there could well be other solutions.