Dr. Z.'s Probability Lecture 6 Handout: Random Variables

By Doron Zeilberger

Important Concept: Random Variable

A **random variable** is really a **function** defined on the **sample space**.

Empirical Examples: 1. The population of people in a county, and $X(person) = HeightOfPerson$.

2. The population of people in a class, and $X(person) = WeightOfPerson$.

3. The population of people in a city, and $X(person) = IQOfPerson$.

Important Concept : Probability mass function for a Discrete Random Variable

If the range of X is "countable" (either finitely many values, or, for example integers or rational numbers), then the *probability mass function* of X is defined by

$$p(a) = P\{X = a\} \quad .$$

In other words, the probability that the value of X happens to be a.

Problem 6.1: In a certain class, the ages are as follows

Abe: 19 ; Barbara: 20; Courtney: 18; David: 20; Evan: 20; Fiona: 19; George: 22; Heather: 20; Igor: 24; Jane: 19;

Find the probability mass function, assuming each person is equally likely to be picked.

Sol. to 6.1: There are 10 students. The ages that show up are $A = \{18, 19, 20, 22, 24\}$.

The set of 18-year-olds is $\{Courtney\}$, and it has one element, hence $p(18) = P\{X = 18\} = \frac{1}{10}$.

The set of 19-year-olds is $\{Abe, Fiona, Jane\}$, and it has three elements, hence $p(19) = P\{X = 19\} = \frac{3}{10}$.

The set of 20-year-olds is $\{Barbara, David, Evan, Heather\}$, and it has four elements, hence $p(20) = P\{X = 20\} = \frac{4}{10} = \frac{2}{5}$.

The set of 22-year-olds is $\{George\}$, and it has one element, hence $p(22) = P\{X = 22\} = \frac{1}{10}$.

The set of 24-year-olds is $\{Igor\}$, and it has one element, hence $p(24) = P\{X = 22\} = \frac{1}{10}$.

For any $a \notin A$, $p(a) = 0$.

Ans. to 6.1: $p(18) = 0.1$, $p(19) = 0.3$, $p(20) = 0.4$, $p(22) = 0.1$, $p(24) = 0.1$, $p(a) = 0$ if $a \notin A$.

Problem 6.2: Let the sample space be the set of all sequences of length n of outcomes of tossing a coin n times, whose probability of a Head is p, and let X be the random variable "Number of Heads". Find the probability mass function of X.

Sol. to 6.2: The range of X is the set $\{0, 1, \ldots, n\}$. We know that the probability of $X = k$ is $\binom{n}{k} p^k (1-p)^{n-k}$, hence

$$p(k) = P\{X = k\} \ = \binom{n}{k} p^k (1-p)^{n-k} \quad .$$

Problem 6.3: The probability mass function of random variable X is given by $c/2^i$, $i = 0, 1, 2, \ldots$, where c is some positive value. Find (i)$=P\{X = 0\}$, (ii)$P\{X < 3\}$, (iii)$= P\{X > 5\}$.

Sol. to 6.3: First we must find c. Since the probabilities have to add-up to 1, we have

$$\sum_{i=0}^{\infty} \frac{c}{2^i} \ = 1 \quad .$$

So

$$c \left(\sum_{i=0}^{\infty} \frac{1}{2^i} \right) \ = 1 \quad .$$

But $\sum_{i=0}^{\infty} \frac{1}{2^i} = 1/(1 - 1/2) = 1/(1/2) = 2$ (recall that if $|x| < 1$, then $1 + x + x^2 + \ldots = 1/(1 - x)$). so $c \cdot 2 = 1$, and we get $c = \frac{1}{2}$, and $P\{X = i\} = \frac{1}{2^{i+1}}$.

Hence

$$P\{X = 0\} = \frac{1}{2^{0+1}} = \frac{1}{2} \quad .$$

$$P\{X < 3\} = P\{X = 0\} + P\{X = 1\} + P\{X = 2\} = \frac{1}{2} + \frac{1}{2^2} + \frac{1}{2^3} = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} = \frac{7}{8} \quad .$$

$$P\{X > 5\} = \sum_{i=6}^{\infty} \frac{1}{2^{i+1}} = \frac{1}{2^7} + \frac{1}{2^8} + \frac{1}{2^9} + \ldots = \frac{1}{2^7} \cdot (1 + \frac{1}{2} + \frac{1}{2^2} + \ldots) = \frac{1}{2^7} \cdot 1/(1 - \frac{1}{2}) = \frac{1}{2^6} = \frac{1}{64} \quad .$$

Important Concept (Expectation of a Discrete Random Variable)

If X is a random variable taking a finite set of values on a finite sample space

$$E[X] \ = \sum_x x p(x) \quad .$$

If X is a random variable taking a discrete set of values

$$E[X] \ = \sum_{x, p(x) > 0} x p(x) \quad .$$

Note: If you have a finite sample space with the uniform distribution the expectation is the simple average.

Problem 6.4: For the random variable of Problem 6.1 find $E[X]$ in two different ways. (i) the simple average (ii) the formula

Sol. to 6.4(i):

$$E[X] = \frac{19 + 20 + 18 + 20 + 20 + 19 + 22 + 20 + 24 + 19}{10} = \frac{201}{10} = 20.1 \quad .$$

Sol. to 6.4(ii):

$$E[X] = \sum_{x \in A} xp(x) = 18 \cdot p(18) + 19 \cdot p(19) + 20 \cdot p(20) + 22 \cdot p(22) + 24 \cdot p(24)$$

$$= 18 \cdot 0.1 + 19 \cdot 0.3 + 20 \cdot 0.4 + 22 \cdot 0.1 + 24 \cdot 0.1 = 20.1 \quad .$$

Problem 6.5: For the random variable of Problem 6.2 find $E[X]$.

Sol. of 6.5:

$$E[X] = \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k} \quad .$$

Since $\binom{n}{k} = \frac{n!}{k!(n-k)!}$, we have

$$k \binom{n}{k} = \frac{k\, n!}{k!(n-k)!} = \frac{n!}{(k-1)!(n-k)!} = n \frac{(n-1)!}{(k-1)!(n-k)!} = n \binom{n-1}{k-1} \quad .$$

Hence (since the summand is 0 when $k = 0$)

$$E[X] = \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=1}^{n} k \binom{n}{k} p^k (1-p)^{n-k} = n \sum_{k=1}^{n} \binom{n-1}{k-1} p^k (1-p)^{n-k}$$

$$= np \sum_{k=1}^{n} \binom{n-1}{k-1} p^{k-1} (1-p)^{n-k} = np \sum_{k-1=0}^{n-1} \binom{n-1}{k-1} p^{k-1} (1-p)^{(n-1)-(k-1)}$$

$$= np \sum_{k=0}^{n-1} \binom{n-1}{k} p^k (1-p)^{n-1-k} = np \cdot (p + (1-p))^n = np \quad .$$

Ans. to 6.5: if X is the Binomial distribution X, with $P(Head) = p$, then $E[X] = np$.

Notes: 1. Later on we will find a simpler way to arrive at this result, using the **linearity of expectation**.

2. **Yet Another way**: Using Calculus, the sum can be written as $((px + (1-p))^n)'$ evaluated at $x = 1$.

Problem 6.6: For the random variable of Problem 6.3 find $E[X]$.

Sol. to 6.6: Since$P\{X=i\} = \frac{1}{2^{i+1}}$, we have

$$E[X] = \sum_{i=0}^{\infty} i \cdot \frac{1}{2^{i+1}} \quad .$$

From calculus we know that

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

(This follows from differentiating $\sum_{i=0}^{\infty} x^i = (1-x)^{-1}$ with respect to x.) Hence

$$E[X] \;=\; \sum_{i=0}^{\infty} i \cdot \frac{1}{2^{i+1}} = \frac{1}{2^2}\left(\sum_{i=0}^{\infty} i \cdot \frac{1}{2^{i-1}}\right) \;=\; \frac{1}{4} \cdot \frac{1}{(1-1/2)^2} = \frac{1}{4} \cdot \frac{1}{1/4} = 1 \quad .$$

Ans. to 6.6: $E[X] = 1$.

Problem 6.7: Two fair three-faced dice are rolled, with faces showing either 1, 2, or 3. Let X equal the product of the number of dots that show up.

(a) Compute $P\{X=i\}$ for $1 \le i \le 9$

(b) Find the expectation $E[X]$.

Sol. to 6.7: (a) The set of possibilities is

$$[1,1],[1,2],[1,3],[2,1],[2,2],[2,3],[3,1],[3,2],[3,3] \quad .$$

The set of products that show up is $A = \{1,2,3,4,6,9\}$.

$$P\{X=1\} = \frac{1}{9} \quad , \quad P\{X=2\} = \frac{2}{9} \quad , \quad P\{X=3\} = \frac{2}{9} \quad ,$$

$$P\{X=4\} = \frac{1}{9} \quad , \quad P\{X=6\} = \frac{2}{9} \quad , \quad P\{X=9\} = \frac{1}{9} \quad .$$

(b)

$$E[X] = 1 \cdot \frac{1}{9} + 2 \cdot \frac{2}{9} + 3 \cdot \frac{2}{9} + 4 \cdot \frac{1}{9} + 6 \cdot \frac{2}{9} + 9 \cdot \frac{1}{9} = 4 \quad .$$

Note: A quicker way would be to do the simple average

$$\frac{(1 \cdot 1 + 1 \cdot 2 + 1 \cdot 3 + \ldots + 3 \cdot 1 + 3 \cdot 2 + 3 \cdot 3)}{9} = \frac{(1+2+3)\cdot(1+2+3)}{9} = \frac{6^2}{9} = 4 \quad .$$

Problem 6.8 Let X be the winnings of a gambler and assume that

$$P\{X=-4\} = 0.25 \quad , \quad P\{X=-3\} = 0.05 \quad , \quad P\{X=-2\} = 0.03 \quad , \quad P\{X=-1\} = 0.07 \quad ,$$

$$P\{X=0\}=0.15 \quad , \quad P\{X=1\}=0.05 \quad , \quad P\{X=2\}=0.17 \quad ,$$

$$P\{X=3\}=0.03 \quad , \quad P\{X=4\}=0.2 \quad .$$

(a) Compute the conditional probability that gambler wins i, for $i=1,2,3$, given that he wins a positive amount.

(b) Find $E[X]$, his expected winning.

Sol. of 6.8(a):

$$P\{X>0\} \;=\; P(X=1)+P(X=2)+P(X=3)+P(X=4) \;=\; 0.05+0.17+0.03+0.2=0.45 \quad .$$

Hence

$$P\{X=1|X>0\}=\frac{P\{X=1\}}{P\{X>0\}}=\frac{0.15}{0.45}=\frac{1}{3} \quad ,$$

$$P\{X=2|X>0\}=\frac{P\{X=2\}}{P\{X>0\}}=\frac{0.17}{0.45}=\frac{17}{45} \quad ,$$

$$P\{X=3|X>0\}=\frac{P\{X=3\}}{P\{X>0\}}=\frac{0.03}{0.45}=\frac{1}{15} \quad ,$$

$$P\{X=4|X>0\}=\frac{P\{X=4\}}{P\{X>0\}}=\frac{0.2}{0.45}=\frac{4}{9} \quad .$$

Sol. of 6.8(b):

$$E[X]=0.25\cdot(-4)+0.05\cdot(-3)+0.03\cdot(-2)+0.07\cdot(-1)$$

$$0.15\cdot(0)+0.05\cdot(1)+0.17\cdot(2)+0.2\cdot(4)$$

$$=-1-0.15-0.06-0.07+0+0.05+0.34+0.8=-0.09 \quad .$$

Ans. to (b): The expected "gain" of the gambler is -0.09. (This is the case in all casinos, since the casino makes money.)

Problem 6.9 The number of injury claims per month is modeled by a random variable N with

$$P[N=n]=\frac{4}{3(n+1)(n+3)}, \quad where \quad n\geq 0 \quad .$$

Determine the probability of at least two claims during a particular month, given that there have been at most four claims during that month.

Sol. to 6.9:

$$P\{N\geq 2|N\leq 4\}=\frac{P\{2\leq N\leq 4\}}{P\{N\leq 4\}}=$$

$$\frac{P\{N=2\}+P\{N=3\}+P\{N=4\}}{P\{N=0\}+P\{N=1\}+P\{N=2\}+P\{N=3\}+P\{N=4\}}$$

$$\frac{4/45+1/18+4/105}{4/9+1/6+4/45+1/18+4/105}=\frac{23}{100} \quad .$$

Ans. to 6.9: the probability of at least two claims during a particular month, given that there have been at most four claims during that month is 0.23.