# Dr. Z.'s Probability Lecture 10 Handout:
# The Geometric, Negative-Binomial and Hypergeometric Distributions; Linearity of Expectation

By Doron Zeilberger

**Version of Dec. 17, 2017**: Thanks to Bennet Greenberg.

**Previous Version of Oct. 17, 2017** (Thanks to Mahima Sharma)

**Pervious version (Oct. 9, 2017)** (Thanks to Rowen Kanj, Samuel Birrer and Clifford Imhomoh.)

**Important Discrete Random Variable (Geometric Random Variable)**

$X$ is a **Geometric** Random Variable with **parameter** $p$ if its *probability mass function* is given by

$$P(X = n) \; = \; (1-p)^{n-1}p, \quad n = 1, 2, \dots .$$

**Note**: W.C. Fields said: *"If you try and fail, try and try again ... then quit, no use being a damn fool about it"*. Well he was wrong! If your probability of a success in one try is $p$, and all the tries are independent, then sooner or later you will succeed. The probability that you will succeed at the $n$-th try (after $n-1$ failures) is $(1-p)^{n-1}p$.

**Note:** If you add-up all the $P(X = n)$, you get a *geometric series*

$$\sum_{n=1}^{\infty}(1-p)^{n-1}p = p \cdot (1 + (1-p) + (1-p)^2 + \dots) \quad ,$$

hence the name! Since $0 < p < 1$ we have $0 < 1 - p < 1$, and we can use the famous formula $1 + x + x^2 + x^3 + \dots = \frac{1}{1-x}$, to evaluate it getting

$$p \cdot \frac{1}{1-(1-p)} = p \cdot \frac{1}{p} = 1 \quad .$$

So the probabilities add-up to 1 and they should!

**Important Formulas**: If $X$ is a geometric random variable with parameter $p$, then its **expectation**, $E[X]$ , and **variance** are given by

$$E[X] \; = \; \frac{1}{p} \quad ; \quad Var(X) \; = \; \frac{1-p}{p^2} \quad .$$

**Problem 10.1**: In a certain try-outs for a basketball varsity team, one makes it to the team if she can score a basket after at most three tries.

(i) If Ellen shoots a basket with probability 0.4, and the outcome of each shot is independent of the others, what is the probability that she would make it to the team?

(ii) What is the expected number of tries until she scores a basket? What is the variance of the number of tries?

**Sol. to 10.1(i)**:

$P(1 \le X \le 3) = P(X=1)+P(X=2)+P(X=3) = 0.6^0 \cdot 0.4 + 0.6^1 \cdot 0.4 + 0.6^2 \cdot 0.4 = (1+0.6+0.36) \cdot 0.4 = 0.784$ .

**Ans. to 10.1(i)**: Ellen's probability of making it to the team is %78.4 .

**Sol. to 10.1(ii)**: $E[X] = \frac{1}{0.4} = 2.5$, $Var(X) = \frac{1-0.4}{0.4^2} = 3.75$, hence $SD(X) = \sqrt{3.75} = 1.93649\ldots$.

**Ans. to 10.1(ii)**: The expected number of tries until Ellen shoots a basket is 2.5 and the standard deviation is $1.93649\ldots$.

**Important Discrete Random Variable (Negative Binomial Random Variable)**

$X$ is a **Negative Binomial** Random Variable with **parameters** $p$ and $r$ if its *probability mass function* is given by

$$P(X=n) = \binom{n-1}{r-1}(1-p)^{n-r}p^r, \quad n = r, r+1, r+2\ldots.$$

It tells you the probability of getting the $r^{th}$ success at the $n^{th}$ try, when the probability of success in a single try is $p$, and all tries are independent.

**Note:** The special case $r = 1$ is a Geometric random variable.

**Important Formulas**: If $X$ is a negative binomial random variable with parameters $p$ and $r$, then its **expectation**, $E[X]$ , and *variance* are given by

$$E[X] = \frac{r}{p} \quad ; \quad Var(X) = \frac{r(1-p)}{p^2} \quad .$$

**Problem 10.2**: In a certain try-outs for a basketball varsity team, one makes it to the team if he or she can score three baskets after at most five tries.

(i) If Tom shoots a basket with probability 0.7, and the outcome of each shot is independent of the others, what is the probability that he would make it to the team?

(ii) What is the expected number of tries until he scores three baskets? What is the variance of the number of tries?

**Sol. to 10.2(i)**:

$$P(3 \le X \le 5) = P(X=3) + P(X=4) + P(X=5)$$

$$= \binom{2}{2}0.3^0 \cdot 0.7^3 + \binom{3}{2}0.3^1 \cdot 0.7^3 + \binom{4}{2}0.3^2 \cdot 0.7^3 = 0.83692 \quad .$$

**Ans. to 10.2(i)**: Tom's probability of making it to the team is %83.692 .

**Sol. to 10.2(ii)**: $E[X] = \frac{3}{0.7} = 4.285714286\ldots$, $Var(X) = \frac{3(1-0.7)}{0.7^2} = 1.836734694\ldots$, hence $SD(X) = \sqrt{1.836734694} = 1.355261854\ldots$.

**Important Discrete Random Variable (Hypergeometric)**

A random variable $X$ is a **Hypergeometric random variable** with parameters $(m, n, N)$ if its probability mass function is

$$P(X = i) = \frac{\binom{m}{i}\binom{N-m}{n-i}}{\binom{N}{n}} \quad , \quad i = 0, 1, 2, \ldots, m \quad .$$

**Note**: Unlike the geometric or negative binomial random variables that can take infinitely many values, a hypergeometric random variable can only take finitely many (in fact $m + 1$ of them).

**Note**: It is used for *quality control*. Suppose you have to pick a sample of size $n$ (where $n$ is small) out of a big batch of $N$ items, of which $m$ items are defective and the remaining $N - m$ are OK. $P(X = i)$ is the probability that our sample will contain exactly $i$ defective items.

**Important Formula**: If $X$ is is a **Hypergeometric random variable** with parameters $(m, n, N)$ then

$$E[X] = \frac{mn}{N} \quad .$$

**Problem 10.3** (corrected, Oct. 17, 2017, thanks to Mahima Sharma): Suppose that a batch of 200 light bulbs contains 10 that are defective and 190 that are OK. If $X$ is the number of defective items in a randomly drawn sample of 20 items from the batch, find $P(X = 0)$ and $P(X < 3)$. Also find the expected number of defective light bulbs in the sample.

**Sol. to 10.3**: Here $N = 200, m = 10, n = 20$. So

$$P(X = i) \ = \ \frac{\binom{10}{i}\binom{190}{20-i}}{\binom{200}{20}} \quad , \quad i = 0, 1, 2, \ldots, 10 \quad .$$

Hence

$$P(X = 0) = \frac{\binom{10}{0}\binom{190}{20}}{\binom{200}{20}} = 0.3397743762\ldots$$

Also

$$P(X < 3) = P(X = 0) + P(X = 1) + P(X = 2) = \frac{\binom{10}{0}\binom{190}{20}}{\binom{200}{20}} + \frac{\binom{10}{1}\binom{190}{19}}{\binom{200}{20}} + \frac{\binom{10}{2}\binom{190}{18}}{\binom{200}{20}}$$

$$= 0.9347145495\ldots \quad .$$

Finally the expected number of defective light bulbs is $\frac{mn}{N} = \frac{10\cdot 20}{200} = 1$.

**Ans. to 10.3**: The probability that the sample of 20 will contain no defective light bulbs is $0.3397743762\ldots$ and that it will contain less than three of them is $0.9347145495\ldots$. The expected number of defective light bulbs is 1.

**Very Important Fact (Linearity of Expectation)**:

If $X$ and $Y$ are random variables defined on the same probability space, then $E[X+Y] = E[X] + E[Y]$. Also, of course, if $c$ is any real number $E[cX] = cE[X]$, and $E[c] = c$. Hence , for any real numbers $a$, $b$ and $c$, we have:

$$E[aX + bY + c] = aE[X] + bE[Y] + c \quad .$$

**Problem 10.4**: In a certain country the average income of the husbands is 1000 Euros per month, the average income of the wives is 1100 per month, and the average income of a child is 100 Euros per month. If the average family has 2.1 children, what is the average family income?

**Sol. to 10.4**: $1000 + 1100 + 2.1 \cdot 100 = 2310$ Euros per month.

**Problem 10.5**: You toss a coin whose probability of Heads is 0.3 1000 times. Let $X$ be the random variable "Twice the Number of Heads in the first 500 tosses Plus three times the Number of Tails in the last 500 tosses". Find $E[X]$.

**Sol. to 10.5**: Let $Y$ be the random variable "Number of Heads in the first 500 tosses" and let $Z$ be the random variable "Number of Tails in the last 500 tosses".

$E[Y] = 500 \cdot 0.3 = 150$, $E[Z] = 500 \cdot 0.7 = 350$, hence $E[X] = E[2Y + 3Z] = 2E[Y] + 3E[Z] = 2 \cdot 150 + 3 \cdot 350 = 1350$.

**Ans. to 10.5**: The expectation of $X$ is 1350.

**Problem 10.6**: You toss coins (of possibly different biases) an even number, $n$, of times.

• Let $Y$ be the random variable "Twice the Number of Heads in the first $n/2$ tosses Plus three times the Number of Heads in the last $n/2$ tosses".

• Let $Z$ be the random variable "The Number of Heads in the first $n/2$ tosses Plus two times the Number of Heads in the last $n/2$ tosses".

If $E[Y] = 130$ and $E[Z] = 80$, what is the expected total number of Heads?

**Sol. to 10.6**: Let $X_1$ be the random variable "Number of Heads in the first $n/2$ tosses" and let $X_2$ be the random variable "Number of Heads in the last $n/2$ tosses".

We have

$$Y \;=\; 2X_1 + 3X_2 \quad , \quad Z \;=\; X_1 + 2X_2 \quad .$$

Taking expectations we have

$$E[Y] = 2E[X_1] + 3E[X_2] \quad , \quad E[Z] = E[X_1] + 2E[X_2] \quad .$$

We have to solve the system of equations

$$2E[X_1] + 3E[X_2] = 130 \quad , \quad E[X_1] + 2E[X_2] = 80 \quad ,$$

that yield $E[X_1] = 20$, $E[X_2] = 30$. The random variable "Total Number of Heads" is $X_1 + X_2$, so, by *linearity of expectation*, we have:

$$E[X_1 + X_2] = E[X_1] + E[X_2] = 20 + 30 = 50 \quad .$$

**Ans. to 10.6**: The expected total number of Heads is 50.

**Comment**: In the original version the coin was the same. In that case, as pointed out by Bennet Greenberg, the data was inconsistent, since the expected number of Heads in the first half is the same as in the second half.