Math 233: Theory of Probability Notes: October 21^{st}

Oishika Chaudhury   Xinbo Li

October 27, 2020

Contents

1 Introduction

This paper covers all the topics that were covered on the 21st October, for Professor Mike O Neil's Theory of Probability class.

2 Other Discrete Probability Distributions

2.1 The geometric random variable

We suppose that independent trials, each with probability p, where $0 < p < 1$, of success are performed until a success occurs. If we let X equal the number of trials that are required to do this then,

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

The aforementioned equation holds true because, in order for X to equal n, it is necessary and sufficient that the first $n-1$ trials are failures and the n^{th} trial is a success. The equation then follows since the outcomes of the successive trials are assumed to be independent.

Since:

$$\sum_{n=1}^{\infty} P(X = n) = p\sum_{n=1}^{\infty}(1-p)^{n-1} = \frac{p}{1-(1-p)} = 1$$

It follows that, a success will eventually occur with probability 1. Any random variable X, whose probability mass function is given by:
$P(X = n) = (1-p)^{n-1}p$
is said to be a geometric random variable with parameter p.

Question An urn contains N white and M black balls. Balls are randomly selectedm one at a time, until a black one is obtained. If we assume that each ball selected is replaced before the next one is drawn, what is the probability that:
(a) exactly n draws are needed?
(b) at least k draws are needed?

Solution If we let X be the number of draws needed to select a black ball, then X satisfies the probability mass function as given by the geometric random variable above with $p = \frac{M}{M+N}$.

(a) $P(X = n) = (\frac{N}{M+N})^{n-1}\frac{M}{M+N} = \frac{MN^{n-1}}{(M+N)^n}$
(b)

$$P(X \geq k) = \frac{M}{M+N}\sum_{n=k}^{\infty}(\frac{N}{M+N})^{n-1}$$

$$= \frac{\frac{M}{M+N}(\frac{N}{M+N})^{k-1}}{1-\frac{N}{M+N}}$$

$$= (\frac{N}{M+N})^{k-1}$$

Question Find the expected value and variance of a geometric random variable.

Solution Let's take $q = 1 - p$, we have:

$$E[X] = \sum_{i=1}^{\infty} iq^{i-1}p$$

=

$$\sum_{i=1}^{\infty}(i-1+1)q^{i-1}p$$

=

$$\sum_{i=1}^{\infty}(i-1)q^{i-1}p + \sum_{i=1}^{\infty} q^{i-1}p$$

=

$$\sum_{j=0}^{\infty} jq^{j}p + 1$$

=

$$q\sum_{j=1}^{\infty} jq^{j-1}p + 1$$

=

$$qE[X] + 1$$

giving us:

$$pE[X] = 1$$

$$E[X] = \frac{1}{p}$$

To determine Var(X), we first compute $E[X^2]$. With $q = 1 - p$, we have that:

$$E[X^2] = \sum_{i=1}^{\infty} i^2 q^{i-1}p$$

=

$$\sum_{i=1}^{\infty}(i-1+1)^2 q^{i-1}p$$

=

$$\sum_{i=1}^{\infty}(i-1)^2 q^{i-1}p + \sum_{i=1}^{\infty} q^{i-1}p + \sum_{i=1}^{\infty} 2(i-1)q^{i-1}p$$

=

$$\sum_{j=0}^{\infty} j^2 q^j p + 2\sum_{j=1}^{\infty} j q^j p + 1$$

=

$$qE[X^2] + 2qE[X] + 1$$

Using, $E[X] = \frac{1}{p}$, the equation for $E[X^2]$ yields,

$$pE[X^2] = \frac{2q}{p} + 1$$

Hence,

$$E[X^2] = \frac{2q+p}{p^2} = \frac{q+1}{p^2}$$

giving the result,

$$Var(X) = \frac{q+1}{p^2} - \frac{1}{p^2} = \frac{q}{p^2} = \frac{1-p}{p^2}$$

2.2 The negative binomial random variable

Suppose that independent trials, each having probability p, $0 < p < 1$, of being a success, are performed until a total of r successes is accumulated. If we let X equal the number of trials required, then:

$$P(X = n) = \binom{n-1}{r-1} p^r (1-p)^{n-r}$$

where

$$n = r, r+1...$$

The aforementioned equation follows because, in order for success to occur at the n^{th} trial, there must be $r-1$ successes in the first $n-1$ trials and the n^{th} trial must be a success. The probability of the first event is:

$$\binom{n-1}{r-1}p^{r-1}(1-p)^{n-r}$$

and the probability of the second is p; thus by independence, the given equation is established. To verify that a total of r successes must eventually be accumulated, we can prove analytically that:

$$\sum_{n=r}^{\infty} P(X=n) = \sum_{n=r}^{\infty} \binom{n-1}{r-1}p^{r}(1-p)^{n-r} = 1$$

Question The Banach match problem: At all times, a pipe-smoking mathematician carries 2 matchboxes- 1 in his left-hand pocket and one in his right-hand pocket. Each time he needs a match, he is equally likely to take it from either pocket. Consider the moment when the mathematician first discovers that one of his matchboxes is empty. If it is assumed that both matchboxes initially contained N matches, what is the probability that there are exactly k matches, k = 0,1,....N, in the other box?

Solution Let E denote the event that the mathematician discovers his right-hand matchbox is empty and there are k matches left in his left-hand pocket at the time. Now, this event will occur if and only if the (N+1) choices of the right-hand matchbox is made at the $(N+1+N-k)^{th}$ trial. Therefore we see from the preceeding equation that:

$$P(E) = \binom{2N-k}{N}\frac{1}{2}^{(2N-k+1)}$$

Since there is an equal probability that the left-hand box is first discovered to be empty, the final probability is twice the value given above,and is given by:

$$P(E) = \binom{2N-k}{N}\frac{1}{2}^{(2N-k)}$$

Question Give the expected value and the variance of a negative binomial. random variable with parameters r and p.

Solution We have:

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

since $n\binom{n-1}{r-1} = r\binom{n}{r}$

$$= \frac{r}{p} \sum_{n=r}^{\infty} n^{k-1} \binom{n}{r} p^{r+1} (1-p)^{n-r}$$

by setting m = n+1

$$= \frac{r}{p} \sum_{m=r}^{\infty} (m-1)^{k-1} \binom{m-1}{r} p^{r+1} (1-p)^{m-(r+1)}$$

$$= \frac{r}{p} E[(y-1)^{k-1}]$$

where Y is a negative binomial random variable with parameters, r+1, p. Setting k = 1 in the preceding equation yields:

$$E[X] = \frac{r}{p}$$

Setting k = 2 in the equation for $E[X^k]$ and using the formula for the expected value of a negative binomial random variable gives:

$$E[X^2] = \frac{r}{p} E[Y-1]$$

$$\frac{r}{p}(\frac{r+1}{p} - 1)$$

Therefore,

$$Var(X) = \frac{r}{p}(\frac{r+1}{p} - 1) - (\frac{r}{p})^2$$

$$= r\frac{1-p}{p^2}$$

2.3 The hypergeometric random variable

Suppose that a sample of size n is to be randomly chosen(without replacement) from an urn containing N balls, of which m are white, and N-m are black. If we let X denote the number of white balls selected then:

$$P(X = i) = \frac{\binom{m}{i}\binom{N-m}{n-i}}{\binom{N}{n}}$$

for i = 0,1,2,...n

A random variable X whose probability mass function is given by the above for some values of n,N,m are said to be a hypergeometric random variable.

Example An unknown number, say N, of animals inhabit a certain region. To obtain some information about the size of the population, ecologists often perform the following experiment: They first catch a number, say m, of these animals, mark them in some manner and release them. After allowing the marked animals to disperse throughout the region, a new catch of size, say n, is made. Let X denote the number of marked animals in the second capture. If we assume that the population of the animals in the region remain fixed between the time of the two catches and that each time an animal was caught it was equally likely to be any of the remaining uncaught animals, it follows that X is a hypergeometric variable such that:

$$P(X = i) = \frac{\binom{m}{i}\binom{N-m}{n-i}}{\binom{N}{n}} = P_i(N)$$

Suppose now that X is observed to equal i. Then, since $P_i(N)$ represents the probability of an observed event when there are actually N animals present in the region, it would appear that a reasonable estimate of N would be the value of N that maximises $P_i(N)$. Such an estimate is called the Maximum Likelihood Estimate. The maximisation of $P_i(N)$ can be done most simply by first noting that:

$$\frac{P_i(N)}{P_i(N-1)} = \frac{(N-m)(N-n)}{N(N-m-n+i)}$$

Now the preceding ratio is greater than 1 if and only if,

$$(N-m)(N-n) \geq N(N-m-n+i)$$

or equivalently if:

$$N \leq \frac{mn}{i}$$

Thus $P_i(N)$ is first increasing and then decreasing and reaches its maximum at the largest integral value not exceeding $\frac{mn}{i}$. This value is the maximum likelihood for N.

Question Determine the expected value and the variance of X, a hypergeometric random variable with parameters n, N and m.

Solution

$$E[X^k] = \sum_{i=0}^{n} i^k P(X=i)$$

$$= \sum_{i=1}^{n} \frac{i^k \binom{m}{i}\binom{N-m}{n-i}}{Nn}$$

Using the identities,

$$i\binom{m}{i} = m\binom{m-1}{i-1}$$

and

$$n\binom{N}{n} = N\binom{N-1}{n-1}$$

, we obtain:

$$E[X^k] = \frac{mn}{N}\sum_{i=1}^{n} i^{k-1} \frac{\binom{m-1}{i-1}\binom{N-m}{n-i}}{\binom{N-1}{n-1}}$$

$$= \frac{mn}{N}\sum_{j=0}^{n-1} (j+1)^{k-1} \frac{\binom{m-1}{j}\binom{N-m}{n-i-j}}{\binom{N-1}{n-1}}$$

=

$$\frac{nm}{N} E[(Y+1)^{k-1}]$$

where Y is a hypergeometric random variable with parameters n-1, N-1, and m-1. Hence upon setting k = 1, we have:

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

Upon setting k = 2 in the equation for $E[X^k]$, we obtain:

$$E[X^2] = \frac{mn}{N}E[Y+1]$$

$$= \frac{mn}{N}[\frac{(n-1)(m-1)}{N-1} + 1]$$

, where the final equality uses our preceding result to compute the expected value of the hypergeometric variable Y. Because $E[X] = \frac{mn}{N}$, we conclude that:

$$Var(X) = \frac{mn}{N}[\frac{(n-1)(m-1)}{N-1} + 1 - \frac{mn}{N}]$$

Letting $p = \frac{m}{N}$ and using the identity:

$$\frac{m-1}{N-1} = \frac{Np-1}{N-1} = p - \frac{1-p}{N-1}$$

shows that:

$$Var(X) = np[(n-1)p - (n-1)\frac{1-p}{N-1} + 1 - np]$$

$$= np(1-p)(1 - \frac{n-1}{N-1})$$

2.4 Bonus: The Zeta(or Zipf) distribution

A random variable is said to have a zeta(sometimes called Zipf) distribution if its probability mass function is given by:

$$P(X = k) = \frac{C}{k^{\alpha+1}}$$

k = 1,2... for some value $\alpha > 0$. Since the sum of the foregoing probabilities, must equal 1 we have that:

$$C = [\sum_{k=1}^{\infty}(\frac{1}{k})^{\alpha+1}]^{-1}$$

2.5 Expected value of sum of random variables

Proposition 9.1 First we need to prove that $E[X] = \sum_{s \in S} X(s)p(s)$ (Proposition 9.1) given that the distinct values for X are x_i and that for each i, S_i is the event that $X = x_i$, or $S_i = \{s : X(s) = x_i\}$
.

Proof

$$
\begin{aligned}
E[X] &= \sum_i x_i P\{X = x_i\} \\
&= \sum_i x_i P\{S_i\} \\
&= \sum_i x_i \sum_{s \in S_i} p(s) \\
&= \sum_i \sum_{s \in S_i} x_i p(s) \\
&= \sum_i \sum_{s \in S_i} X(s)p(s) \\
&= \sum_{s \in S} X(s)p(s)
\end{aligned}
$$

Note: The final equality follows because the union of mutually exclusive events $S_1, S_2, ...$ is S

This is intuitive because the expected value, which is the weighted average of the possible values of X, each weighted by the probability that X assumes that value, should equal a weighted average of the values $X(s), s \in S$ each weighted by the probability that s is the outcome of the experiment.

Example Now let's consider an example that illustrates the above proposition. Suppose that two coins are flipped independently, with a probability of p that heads come up. Let X denote the number of heads obtained.

$$P(X=0) = P(t,t) = (1-p)^2 \quad (1)$$
$$P(X=1) = P(h,t) + P(t,h) = 2p(1-p) \quad (2)$$
$$P(X=0) = P(h,h) = p^2 \quad (3)$$
$$\quad (4)$$

By definition, $E[X] = 0 \cdot (1-p)^2 + 1 \cdot 2p(1-p) + 2 \cdot^2 = 2p$

Now, we find out if this agrees with the right hand side of the equation of Proposition 9.1:
Since s, the outcome of the experiment,

$$\sum_{s \in S} X(s)p(s) = X(h,h)p^2 + X(t,t)(1-p)^2 + X(h,t)p(1-p) + X(t,h)(1-p)p \quad (5)$$
$$= 2p^2 + 0 + p_p^2 + p - p^2 \quad (6)$$
$$= 2p \quad (7)$$

Corollary 9.2 Now, we can move on and prove that the expected values of a sum of random variables is equal to the sum of their expected values. (Corollary 9.2)
Let X and Y be two random variables.Then $Z = X + Y$ is also a random variable. Let $s \in S$ be an individual outcome in the sample space. Then $Z(s) = X(s) + Y(s)$ We now show that $E[Z] = E[X+Y] = E[X] + E[Y]$.

$$E[Z] = \sum_{s \in S} Z(s)p(s) \quad \text{by Proposition 9.1} \quad (8)$$
$$= \sum_{s \in S} (X(S) + Y(s))p(s) \quad (9)$$
$$= \sum_{s \in S} X(s)p(s) + \sum_{s \in S} Y(s)p(s) \quad (10)$$
$$= E[X] + E[Y] \quad (11)$$

Note that this can be generalized into $E[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} E[X_i]$,

where $Z = \sum_{i=1}^{n} X_i$

Example Consider the expected value of the sum obtained when n fair dice are rolled.
Let X be the sum. We can represent X by

$$X = \sum_{i=1}^{n} X_i$$

where X_i is the value on die i. Since any value of X_i is equally likely,

$$E[X_i] = \sum_{i=1}^{6} i(\frac{1}{6}) = 3.5$$

Thus, by Corollary 9.2,

$$E[X_i] = E[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} E[X_i] = 3.5n$$

Let's verify this with the case where n=2.
With two dice, X can take on any integer from 2 to 12, inclusive. $P(X = 2) = 1/36 \quad P(X = 3) = 2/36$
$P(X = 4) = 3/36 \quad P(X = 5) = 4/36$
$P(X = 6) = 5/36 \quad P(X = 7) = 6/36$
$P(X = 8) = 5/36 \quad P(X = 9) = 4/36$
$P(X = 10) = 3/36 \quad P(X = 11) = 2/36$
$P(X = 12) = 1/36$
Hence,
$E[X_i] = E[\sum_{i=1}^{n} X_i]$
$= \frac{1*0+2*1+3*2+4*3+5*4+6*5+7*6+12*0+11*1+10*2+9*3+8*4}{36} = \frac{252}{36} = 7 = 3.5 \cdot 2$

2.6 Properties of the cumulative distribution function

Recall: For the distribution function F of X, $F(b)$ denotes the probability that the random variable X takes on a value less than or equal to b: $F(x) = P(X \leq b)$
There are four important properties of the cumulative distribution function F:

1. F is a nondecreasing function, meaning that if $a < b$, then $F(a) \leq F(b)$.
2. $\lim_{b \to \infty} F(b) = 1$
3. $\lim_{b \to -\infty} F(b) = 0$
4. F is right continuous, meaning that for any b and any decreading sequence b_n, $n \geq 1$, that converges to b, $\lim_{n \to \infty} F(b_n) = F(b)$

Proofs of properties Property 1 follows because for $a < b$, the event $\{X \leq a\}$ is contained in the event $\{X \leq b\}$. Property 2 follows because if b_n increases to infinity, then the events $\{X \leq b_n\}$ are increasing events whose union is the event $\{X < \infty\}$. Thus, by the continuity property of probabilities,

$$\lim_{n \to \infty} P\{X \leq b_n\} = P\{X < \infty\} = 1$$

Similarly, property 3 follows because if b_n decreases to negative infinity, $\{X \leq b_n\}$ are decreasing events whose intersection is the event $\{X < -\infty\}$.

$$\lim_{n \to \infty} P\{X \leq b_n\} = P\{X < -\infty\} = 0$$

Property 4 follows because if b_n decreases to b, then $\{X \leq b_n\}, n \geq 1$ are decreasing events whose intersection is the event $\{X \leq b\}$. The continuity property yields

$$\lim_{n \to \infty} P\{X \leq b_n\} = P\{X \leq b\}$$

Example Consider the following distribution function of the random variable X:

$$F(x) = \begin{cases} 0 & x < 0, \\ \frac{x}{2} & 0 \leq x < 0.5 \\ \frac{2}{3} & 1 \leq x < 2 \\ \frac{11}{12} & 2 \leq x < 3 \\ 1 & 3 \leq x \end{cases} \tag{12}$$

Compute (a) $P\{X < 3\}$, (b) $P\{X = 1\}$, (c) $P\{X > \frac{1}{2}\}$, and (d) $P\{2 < X \leq 4\}$

(a) $P\{X < 3\} = \lim_{n\to\infty} P\{\leq 3 - \frac{1}{n}\} = \lim_{n\to\infty} F\left(3 - \frac{1}{n}\right) = \frac{11}{12}$

(b)

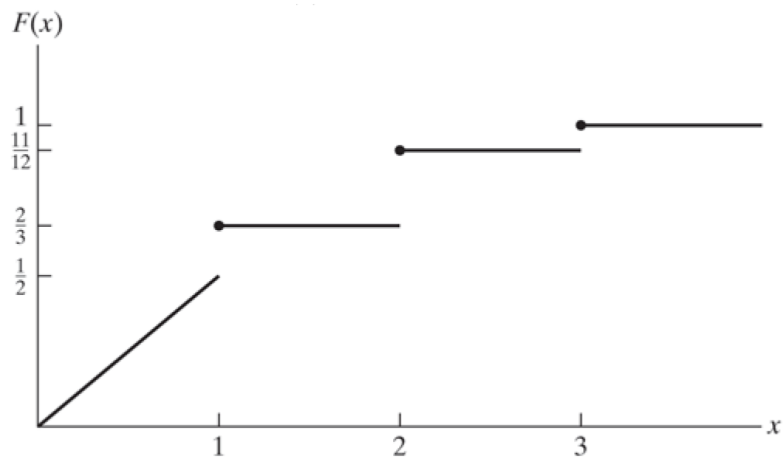
$$P\{X = 1\} = P\{X \leq 1\} - P\{X < 1\} \tag{13}$$
$$= F(1) - \lim_{n\to\infty} F(1 - \frac{1}{n}) = \frac{2}{3} - \frac{1}{2} = \frac{1}{6} \tag{14}$$

(c)

$$P\{X > \frac{1}{2}\} = 1 - P\{X \leq \frac{1}{2} \tag{15}$$
$$= 1 - F(\frac{1}{2}) \tag{16}$$
$$= 1 - \frac{\frac{1}{2}}{2} = \frac{3}{4} \tag{17}$$

(d) $P\{2 < X \leq 4\} = P\{X \leq 4\} - P\{X \leq 2\} = F(4) - F(2) = \frac{1}{12}$.

Below if a graph for $F(x)$.


F(x)
1
11/12
2/3
1/2
1
2
3
x