Random Variables Lecture 6

Binomial Random Variables

n independent trials are performed. Each trial is a success with probability p and failure with probability $1-p$.

Let X be the number of successes. Then $X = 0, 1, \dots, n$ and $P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$ where $\binom{n}{k}$ is the number of choices of the k experiments with successful outcomes and $p^k(1-p)^{n-k}$ is the probability of a particular run of k successes and $n-k$ failures.

For instance if $n=4$ and $k=2$ $P(X=2) =$

1) $\underset{1}{\underline{S}}\ \underset{2}{\underline{S}}\ \underset{3}{\underline{F}}\ \underset{4}{\underline{F}}$

$\left(\frac{1}{3}\right)^2 \cdot \left(\frac{2}{3}\right)^2$ +

2) $\underset{1}{\underline{S}}\ \underset{2}{\underline{F}}\ \underset{3}{\underline{S}}\ \underset{4}{\underline{F}}$

$\frac{1}{3} \cdot \frac{2}{3} \cdot \frac{1}{3} \cdot \frac{2}{3}$

3) $\underset{1}{\underline{S}}\ \underset{2}{\underline{F}}\ \underset{3}{\underline{F}}\ \underset{4}{\underline{S}}$

$\frac{1}{3} \cdot \left(\frac{2}{3}\right)^2 \cdot \frac{1}{3}$ +

4) $\underset{1}{\underline{F}}\ \underset{2}{\underline{S}}\ \underset{3}{\underline{S}}\ \underset{4}{\underline{F}}$

$\frac{2}{3} \cdot \left(\frac{1}{3}\right)^2 \cdot \frac{2}{3}$

5) $\underset{1}{\underline{F}}\ \underset{2}{\underline{S}}\ \underset{3}{\underline{F}}\ \underset{4}{\underline{S}}$

$\frac{2}{3} \cdot \frac{1}{3} \cdot \frac{2}{3} \cdot \frac{1}{3}$ +

6) $\underset{1}{\underline{F}}\ \underset{2}{\underline{F}}\ \underset{3}{\underline{S}}\ \underset{4}{\underline{S}}$

$\left(\frac{2}{3}\right)^2 \cdot \left(\frac{1}{3}\right)^2$

Hence $P(X=2) = \binom{4}{2}\left(\frac{1}{3}\right)^2\left(\frac{2}{3}\right)^2$.

Observe that $\sum_{k=0}^{n} P(X=k) = \sum_{k=0}^{n}\binom{n}{k} p^k (1-p)^{n-k}$

$= (p+(1-p))^n = 1^n = 1$

Ex. Screws produced by a certain manufacturer are defective with probability 0.01 independently of each other. The company sells the screws in packages of 10 and offers money back guarantee that at most 1 of 10 screws are defective. What proportion of packages sold must the company replace?

Solution: Let X = # of defective screws. Then X is a binomial random variable with parameters $(10, 0.01)$

We wish to compute $P(X \geq 2) = 1 - P(X \leq 1) =$

$= 1 - P(X=1) - P(X=0) = 1 - \binom{10}{1} 0.01 \cdot (0.99)^9 - (0.99)^{10}$

≈ 0.004. Thus 4 out of 1000 packages will have to be replaced.

Ex. At each step a randomly moving particle is equally likely to move one unit to the left or one unit to the right on a number line. If the particle executes a total of n moves, describe the probability mass function for the particle's new position given that it starts at 0.

<u>Solution:</u> Let X = # of right moves. Then X is a binomial random variable with parameters $(n, \frac{1}{2})$.

Now, given $X = j$, the particle moved j units to the right and $n-j$ units left for a "total position" of $j-(n-j) = 2j-n$.

Let $Y = 2X - n$ then $P(Y=k) =$

$$= P(2X-n=k) = P\left(X = \frac{n+k}{2}\right) = \begin{cases} \binom{n}{\frac{n+k}{2}}\left(\frac{1}{2}\right)^n & \text{if } n+k \text{ even} \\ 0 & \text{otherwise.} \end{cases}$$

Thus $Y = -n, -n+2, -n+4, -n+8, \ldots, n$.

with $P(Y=-n+2j) = \binom{n}{j}\left(\frac{1}{2}\right)^n$

<u>Ex.</u> In a certain country club there is a popular dish which each guest orders independently with probability 0.8. Today the country club has 100 guests.

(a) What is the probability that all of them will order this dish?

(b) Clearly preparing 100 dishes is wasteful. How many dishes should the country club prepare if management wants the probability of running out of stock to be less than or equal to $\frac{1}{1000}$?

Solution: let $X = \#$ of dishes ordered.

(a) $P(X=100) = (0.8)^{100} \approx 2\times 10^{-10}$.

Very unlikely!

(b) We want to prepare $m < 100$ dishes s.t. the probability $P(X > m) = \sum_{k=m+1}^{100} \binom{100}{k} (0.8)^k (0.2)^{100-k}$ $\leq \frac{1}{1000}$

Experiment shows that this happens for the first time when $m \geq 91$. (We will see how to get to this number more cleverly when we study approximations to the binomial distribution).

Ex. An electronic device has an odd number of identical components. Each components is likely to fail with probability p. The components function independently of one another. The device fails if most of its components fail. What are the conditions on p if the device with $2n+1$ components is more reliable than the device with $2n-1$ components? (E.g. which airplane is safer? one with 3 propellers or one with 5 propellers?)

<u>Solution:</u> Let X be a binomial random variable with parameters $(2n+1, p)$. Let Y be a binomial random variable with parameters $(2n-1, p)$.

Then if we claim that device with more components is more reliable, then

$$\boxed{P(X \leq n) \geq P(Y \leq n-1)}$$

That is, the probability that device X functions is the probability that fewer than $n+1$ components fail. Similarly, the probability that device Y functions is the probability of fewer than n failures.

Let N_0, N_1, N_2 be the events of 0, 1, and 2 failures in the first two components of device X.

Then $P(X \leq n) = P(X \leq n | N_0) p(N_0) + P(X \leq n | N_1) P(N_1)$
$+ p(X \leq n | N_2) p(N_2) = p(Y \leq n)(1-p)^2 + p(Y \leq n-1) \cdot 2p(1-p)$
$+ p(Y \leq n-2) p^2 = \Big(p(Y \leq n-1) + p(Y=n)\Big)(1-p)^2$
$+ p(Y \leq n-1) \cdot 2p(1-p) + \Big(p(Y \leq n-1) - p(Y = n-1)\Big) p^2$
$= p(Y \leq n-1)\left((1-p)^2 + 2p(1-p) + p^2\right) + P(Y=n)(1-p)^2$
$- p(Y=n-1) p^2 = P(Y \leq n-1)\left((1-p)+p\right)^2 + \binom{2n-1}{n} p^n (1-p)^{n+1}$

$$- \binom{2n-1}{n-1} p^{n+1}(1-p)^n = P(Y \leq n-1) +$$

$$+ \binom{2n-1}{n}\left(p^n(1-p)^{n+1} - p^{n+1}(1-p)^n\right) \geqslant P(Y \leq n-1)$$

Thus $p^n(1-p)^{n+1} - p^{n+1}(1-p)^n \geqslant 0$

or $1-p-p = 1-2p \geqslant 0.$

Hence $\frac{1}{2} \geqslant p.$ In particular, a system with more components is more reliable whenever probability of failure is less that $\frac{1}{2}$.

Properties of binomial random variables

We will compute the expected value and variance of a binomial random variable with parameters (n, p) in two ways. The hard way and the easy way. Lets start with the hard way first.

(a) Without using properties of expectation

We need to compute $E[X^1]$ and $E[X^2]$. We might as well find $E[X^k]$ for any integer $k \geqslant 0$.

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

$$i\binom{n}{i} = n\binom{n-1}{i-1}$$ we get.

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

$$= np \sum_{j=0}^{n-1} (j+1)^{k-1} \binom{n-1}{j} p^j (1-p)^{n-1-j} = np\, E[(Y+1)^{k-1}]$$

where Y is a binomial random variable with parameters $(n-1, p)$.

Thus $E[X] = E[X^1] = np\, E[(Y+1)^0] = np\, E[1]$

$= np.$

and $E[X^2] = np\, E[(Y+1)^1] = np(E[Y]+1) =$

$= np([n-1]p+1)$

Hence $Var(X) = E[X^2] - (E[X])^2 = np([n-1]p+1) - (np)^2$

$= np - np^2 = np(1-p)$

(b) Using Properties of expectation

Recall that $X = X_1 + X_2 + \ldots + X_n$ where

$$X_k = \begin{cases} 1 & \text{if } k^{th} \text{ trial is a success.} \\ 0 & \text{otherwise.} \end{cases}$$

Then $E[X] = E[X_1] + E[X_2] + \ldots + E[X_n]$

$\qquad = n\, E[X_1] = np.$

Since the X_k are independent, we also have

$Var(X) = Var(X_1) + Var(X_2) + \ldots + Var(X_n) = n\, Var(X_1)$

Now $E[X_i^2] = p$. Therefore $Var(X_1) = E[X_i^2] - \left(E[X_1]\right)^2$
$= p - p^2 = p(1-p)$.

Hence $Var(X) = n\,Var(X_1) = np(1-p)$.

The following proposition details how the binomial probability mass function first increases and then decreases.

<u>Proposition:</u> If X is a binomial random variable with parameters (n,p) where $0<p<1$, then as k goes from 0 to n, $P(X=k)$ first increases monotonically and then decreases monotonically, reaching its largest value when k is the largest integer less than or equal to $(n+1)p$.

<u>Proof:</u> $P(X=k) \geq P(X=k-1)$ iff $\dfrac{P(X=k)}{P(X=k-1)} \geq 1$

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

$$= \frac{n+1-k}{k} \cdot \frac{p}{1-p} \geq 1 \quad \text{iff}$$

$$np + p - kp \geq k - kp \quad \text{iff} \quad np + p \geq k \quad \text{iff} \quad (n+1)p \geq k.$$

Hypergeometric Random Variables:

Suppose that an urn contains w white and b black balls and we draw a random sample of $n \leq w+b$ balls. Let $X = 0, 1, \dots, n$ be the number of white balls picked.

(a) If the balls are drawn sequentially with replacement X is a binomial random variable with parameters $\left(n, \frac{w}{w+b}\right)$.

(b) If we draw the n balls without replacement X is said to be hypergeometric with parameters (w, b, n).

Notice that $P(X=k) = \dfrac{\binom{w}{k}\binom{b}{n-k}}{\binom{w+b}{n}}$

where $0 \leq k \leq \min\{w, n\}$.

Observe that $\displaystyle\sum_{k=0}^{\min\{w,n\}} P(X=k) = 1$.

For instance if $n \leq w$

$$\sum_{k=0}^{n} P(X=k) = \frac{\sum_{k=0}^{n}\binom{w}{k}\binom{b}{n-k}}{\binom{w+b}{n}} = \frac{\binom{w+b}{n}}{\binom{w+b}{n}} = 1.$$

Ex. A forest has N elk. Today m of the elk are captured, tagged, and released. At a later date, n elk are randomly captured. Assume that each of the N elk is still equally likely to be captured. What is the probability that k of the original elk are recaptured?

Solution: $$P(X=k) = \frac{\binom{m}{k}\binom{N-m}{n-k}}{\binom{N}{n}}.$$

Ex. A five-card hand is drawn at random. The number of aces in the hand is $X = 0, 1, \ldots, 4$.

with $$P(X=k) = \frac{\binom{4}{k}\binom{48}{5-k}}{\binom{52}{5}}.$$

Thus X is a hypergeometric random variable with parameters $(4, 48, 5)$.

Thm: Two hypergeometric random variables with respective parameters (w, b, n) and $(n, w+b-n, w)$ have identical distributions.

Proof: Suppose we have $w+b$ balls. Imagine these balls are initially identical. At a later time n of the balls are stamped with the letters s (for sampled) and w (for white). Then sampling n ~~white~~ balls and checking the number of white balls can be done in two ways

① Stamp "w" on w balls (designating that they are white) then randomly stamp "s" on n balls.

Then

$$P(X=k) = \frac{\binom{w}{k}\binom{b}{n-k}}{\binom{w+b}{n}}$$

where X is the number of white balls in sample.

② Stamp "s" on n balls then randomly stamp "w" on w balls. (i.e. first the balls are sampled then colored).

$$P(X=k) = \frac{\binom{n}{k}\binom{w+b-n}{w-k}}{\binom{w+b}{w}}$$

By symmetry ① = ②.

<u>Properties of hypergeometric random variables</u>

Let X be a hypergeometric random variable with parameters (w, b, n) and $X = \#$ of white balls in sample.

Then $X = X_1 + X_2 + \ldots + X_n$ where $X_k = \begin{cases} 1 & \text{if } k^{th} \text{ ball picked} = \text{"w"} \\ 0 & \text{otherwise.} \end{cases}$

$$E[X] = nE[X_1] = n\frac{w}{w+b} = \frac{nw}{w+b}.$$

To compute variance, observe that unlike the binomial case, the X_k are not independent. For instance, $X_1 = 1$ makes it less likely that $X_2 = 1$ because there are fewer white balls to draw from if the first ball that is removed is white.

$$Var(X) = Var(X_1 + \dots + X_n) = \sum_{k=1}^{n} Var(X_k) +$$

$$+ \sum_{k=1}^{n} \sum_{j \neq k} Cov(X_k, X_j) = n\,Var(X_1) + n(n-1)\,Cov(X_1, X_2)$$

(why?)

Now, $E[X_1^2] = \frac{w}{w+b}$ so $Var(X_1) = \frac{w}{w+b} - \left(\frac{w}{w+b}\right)^2$

$$= \frac{w}{w+b}\left(1 - \frac{w}{w+b}\right) = \frac{w}{w+b} \cdot \frac{b}{w+b}.$$

$E[X_1 X_2] = \frac{w(w-1)}{(w+b)(w+b-1)}$ so $Cov(X_1, X_2) =$

$$= \frac{w(w-1)}{(w+b)(w+b-1)} - \left(\frac{w}{w+b}\right)^2 = \frac{w}{w+b}\left(\frac{-b}{(w+b)(w+b-1)}\right)$$

Thus, $Var(X) = \frac{nwb}{(w+b)^2} - \frac{n(n-1)bw}{(w+b)^2(w+b-1)}$

$$= \frac{nwb}{(w+b)^2}\left(1 - \frac{(n-1)}{w+b-1}\right) = \frac{nwb}{(w+b)^2} \cdot \frac{(w+b-n)}{w+b-1}$$

$$= \frac{nwb(w+b-n)}{(w+b)^2(w+b-1)}.$$