Christopher Lum
lum@uw.edu

Lecture09c
Binomial Distribution

The YouTube video entitled 'TBD' that covers this lecture is located at .

Binomial Distribution

The binomial distribution occurs in games of chance (rolling a die), quality inspection (counting the number of defectives), opinion polls (counting number of employees favoring something), medicine (number of patients who recover from a certain medicine), etc. The conditions of its occurrence are as follows.

We are interested in the number of times an event A occurs in n independent trials. In each trial the event A has the same probability $P(A) = p$. Then in a trial, A will *not* occur with probability $q = 1 - p$. In n trials, the random variable that interests us is

X = Number of times the event A occurs in n trials.

X can assume the values 0, 1, ... , n and we want to determine the corresponding probabilities $(p(X = 0),\ p(X = 1),\ \ldots,\ p(X = n))$.

Note that $X = x$ means that A occurs in x trials an in $n - x$ trials it does not occur.

It may be easiest at this point to consider a concrete example. Consider the case of $n = 5$ and we are interested in $X = 2$. In this case, one such series that may occur is

$$A\ A\ B\ B\ B$$

where $B = A^c$ (meaning A does not occur)

Assuming that the events are independent, we see that the probability of the sequence $A\ A\ B\ B\ B$ is given by

$$P(A)\ P(A)\ P(B)\ P(B)\ P(B) = p\ p\ q\ q\ q = p^2\ q^3$$

Another such sequence which could occur and would lead the $X = 2$ is

$$A\,B\,A\,B\,B$$

By identical reasoning, we obtain the probability of this sequence is also $p^2\,q^3$.

In fact, we see that there are 5 given things (trials) which can be divided into 2 classes of alike things (A or B). Recall that Theorem 2b of section 24.4 described the permutations of classes of equal things (which is the scenario we have here) and we see that we have

$$\binom{n}{k} = \frac{n!}{k!\,(n-k)!} = \frac{5!}{2!\times(5-2)!} = 10 \qquad \text{(“n choose k” , binomial coefficient)}$$

ways that we can obtain $X = 2$. These sequences are enumerated for completeness

1. AABBB
2. ABABB
3. ABBAB
4. ABBBA
5. BABBA
6. BBABA
7. BBBAA
8. BBAAB
9. BAABB
10. BABAB

Since any of these sequences yields $X = 2$, we have

$$P(X = 2) = P(\text{AABBB}) + P(\text{ABABB}) + \ldots + P(\text{BABAB})$$

Recall that each of these have equal probability of $p^2\,q^3$ so we have

$$P(X = 2) = p^2\,q^3 + p^2\,q^3 + \ldots + p^2\,q^3$$

$$= 10\,p^2\,q^3$$

$$P(X = 2) = \frac{5!}{2!\times(5-2)!}\,p^2\,q^3$$

Now that we understand how the concrete example works, we can generalize this logic to a general $P(X = x)$, $x = 1,\ 2,\ \ldots,\ n$. For the general case, we see a sequence that yields $X = x$ must look something like x occurrences of A and therefore $n - x$ occurrences of B. One such sequence is

$$(A\,A\ \ldots\,A)\,(\,B\,B\,\ldots\,B) \qquad \text{where there are } x\ A's\text{, and } n - x \text{ B's} \qquad \textbf{\textit{(Eq.1)}}$$

Therefore, the probability of any of these sequences is

$$(p(A)\,p(A)\,...\,p(A))\,(p(B)\,p(B)\,...\,p(B)) = (p\,p\,...\,p\,)\,(q\,q\;\,...\,q) = p^x\,q^{n-x} \qquad \textbf{(Eq.1*)}$$

Theorem 2b of section 24.4 gives the number of different sequences that yield $X = x$ as the binomial coefficient of n choose x

$$\frac{n!}{x!\,(n-x)!} = \binom{n}{x}$$

Finally, the combined probability of any sequence that yields $X = x$ is obtained by multiplying the binomial coefficient by the probabilities of each individual probability to obtain

$$p(X = x) = f(x) = \begin{cases} \binom{n}{x} p^x\,q^{n-x} & x = 0,\ 1,\ \ldots,\ n\ \text{(integer)} \\ 0 & \text{otherwise} \end{cases} \qquad \textbf{(Eq.2)}$$

where $\quad p = p(A)$

$\qquad\quad q = p(B) = p(A^c) = 1 - p$

The discrete distribution described by Eq.2 is called the **binomial distribution** or **Bernoulli distribution**. Occurrences of A are called success and B are failures.

We can show (in a homework assignment) that the mean of the binomial distribution is

$$\mu = n\,p \qquad \textbf{(Eq.3)}$$

and the variance is given by

$$\sigma^2 = n\,p\,q = n\,p\,(1 - p) \qquad \textbf{(Eq.4)}$$

For the symmetric case of equal chance of success and failure ($p = q = 1/2$), this gives

$$\mu = n/2$$

$$\sigma^2 = n/4$$

$$f(x) = \binom{n}{x}\left(\tfrac{1}{2}\right)^n \qquad (x = 1,\ 2,\ \ldots,\ n)$$

Going back to our example of $n = 5$, we calculated $f(2)$, and can follow the same procedure (Eq.2) to obtain the other values of the probability function. Repeating for different values of p yields Figure 517.

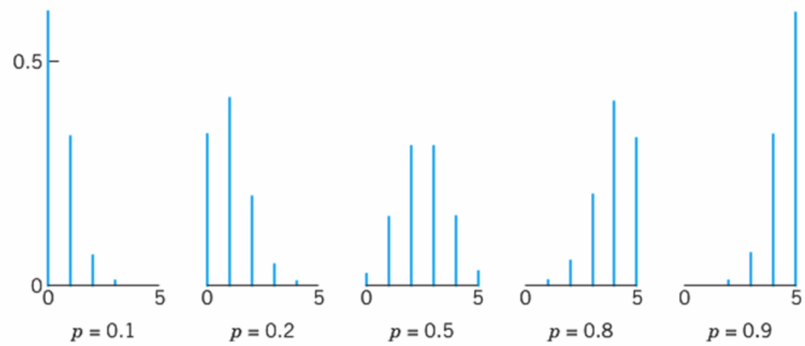


Fig. 517. Probability function (2) of the binomial distribution for $n = 5$ and various values of p

As expected, as the probability of success increases (p increases), the probability of obtaining more successes over the $n = 5$ trials increases.

Example 1: Binomial Distribution

Compute the probability of obtaining <u>at least</u> two six's in rolling a fair die 4 times.

Solution: We consider obtaining a "six" a success. Therefore, we have

$$p = p(A) = p(\text{"six"}) = \frac{1}{6}$$

$$q = p(A^c) = 1 - p(A) = 1 - \frac{1}{6} = \frac{5}{6}$$

If we consider the random variable X = number of sixes that occur from $n = 4$ trials, we can use Eq.2 to compute probability of $X = x$. For example, for $p(X = 2) = f(2)$ is given by

$$p(X = 2) = f(2) = \binom{4}{2} p^2 q^{4-2} = 6 \left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right)^2 = \frac{25}{216} \quad \text{(probability that we obtain exactly 2 sixes)}$$

Similarly for $f(3)$ and $f(4)$, we obtain

$$p(X = 3) = f(3) = \binom{4}{3} p^3 q^{4-3} = 4 \left(\frac{1}{6}\right)^3 \left(\frac{5}{6}\right)^1 = \frac{5}{324} \quad \text{(probability that we obtain exactly 3 sixes)}$$

$$p(X = 4) = f(4) = \binom{4}{4} p^4 q^{4-4} = 1 \left(\frac{1}{6}\right)^4 1 = \frac{1}{1296} \quad \text{(probability that we obtain exactly 4 sixes)}$$

In[1]:=
```
p = 1/6;
q = 5/6;
n = 4;

x = 2;
Print["f(", ToString[x], ") = "]
f2 = Binomial[n, x] p^x q^(n-x)
f2 // N
Print[]

x = 3;
Print["f(", ToString[x], ") = "]
f3 = Binomial[n, x] p^x q^(n-x)
f3 // N
Print[]

x = 4;
Print["f(", ToString[x], ") = "]
f4 = Binomial[n, x] p^x q^(n-x)
f4 // N
Print[]
```

f(2) =

Out[6]= $\frac{25}{216}$

Out[7]= 0.115741

f(3) =

Out[11]= $\frac{5}{324}$

Out[12]= 0.0154321

f(4) =

Out[16]= $\frac{1}{1296}$

Out[17]= 0.000771605

Since we are interested in obtaining <u>at least</u> 2 sixes, this means we want the probability that $X = 2, 3,$ or 4, so the total probability that we are interested in, P is given by

$$P = f(2) + f(3) + f(4) = \frac{171}{1296} = \frac{19}{144} \approx 13.2 \times \%$$

In[19]:=
```
f2 + f3 + f4
f2 + f3 + f4 // N
```

Out[19]= $\frac{19}{144}$

Out[20]= 0.131944

Mathematica provides the 'BinomialDistribution' function to describe this distribution. We can evaluate it at certain points by using it in conjunction with the 'PDF' function

In[21]:=
```
(*Evaluate the distribution at the specified x values*)
PS = PDF[BinomialDistribution[n, p], {0, 1, 2, 3, 4}]
```

Out[21]= $\left\{\frac{625}{1296}, \frac{125}{324}, \frac{25}{216}, \frac{5}{324}, \frac{1}{1296}\right\}$

As can be seen, this is exactly what we obtained previously.

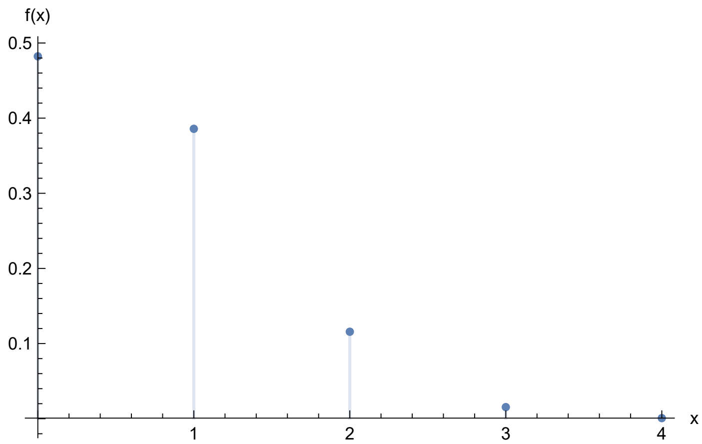
We can visualize this distribution. To do this, we first define a function which returns the appropriate probability

In[22]:=
```
(*Define the probability distribution function*)
f[x_] := Module[
  (*Define local variables*)
  {binIndex},

  (*Function body*)

  (*Check if x is not an integer or if it is outside range [2,12]*)
  If[! Element[x, Integers],
   Return[0],
  ];

  If[x > 4,
   Return[0],
  ];

  If[x < 0,
   Return[0],
  ];

  binIndex = x + 1;
  Return[PS[[binIndex]]]
 ]
```

We can now plot the probability distribution function, $f(x)$, using Mathematica's 'DiscretePlot' function.

In[23]:=
```
xMin = 0;
xMax = 4;
dx = 1;
DiscretePlot[f[x], {x, xMin, xMax, dx},
 AxesLabel → {"x", "f(x)"},
 PlotRange → All]

Clear[xMin, xMax, dx]
```

Out[26]=



In[28]:=
```
Clear[f, PS, f2, f3, f4]
```

Example: Liar’s Dice

Show example from Red Dead Redemption where you can play Liar’s Dice and talk about how this can be modeled with a Binomial distribution.

YouTube - Red Dead Redemption: How to Play Liar’s Dice (link)

In this case, assume that each player has 3 die each.

Let us first consider the scenario where all 6 dice are hidden from both players (for example both players are playing against the house). Suppose that house claims (AKA bids) 2 “fours”, meaning that between the 6 dice, there are at least two die with the face value of four. If we consider the random variable to be the number of “fours” that are in the set, we see that the probability distribution of X is a standard binomial distribution with $n = 6$.

In[29]:=
```
(*binomial distribution parameters if all die are blind*)
n = 6;
p = 1 / 6;

(*Evaluate the distribution at the specified x values*)
PS = PDF[BinomialDistribution[n, p], {0, 1, 2, 3, 4, 5, 6}]
```

Out[31]=

$$\left\{\frac{15625}{46656}, \frac{3125}{7776}, \frac{3125}{15552}, \frac{625}{11664}, \frac{125}{15552}, \frac{5}{7776}, \frac{1}{46656}\right\}$$

So we have

$P(X=0)$	$P(X=1)$	$P(X=2)$	$P(X=3)$	$P(X=4)$	$P(X=5)$	$P(X=6)$
0 successes	1 success	2 success	3 success	4 success	5 success	6 success
$\frac{15625}{46656}$	$\frac{3125}{7776}$	$\frac{3125}{15552}$	$\frac{625}{11664}$	$\frac{125}{15552}$	$\frac{5}{7776}$	$\frac{1}{46656}$

So we see that the total probability of the bid of at least 2 “fours” being true is given by

$$f(2)+f(3)+f(4)+f(5)+f(6)$$

In[32]:=
```
(*sum the probabilities (watch out for the off by one error)*)
pTwo = PS⟦3⟧ + PS⟦4⟧ + PS⟦5⟧ + PS⟦6⟧ + PS⟦7⟧
pTwo // N
```

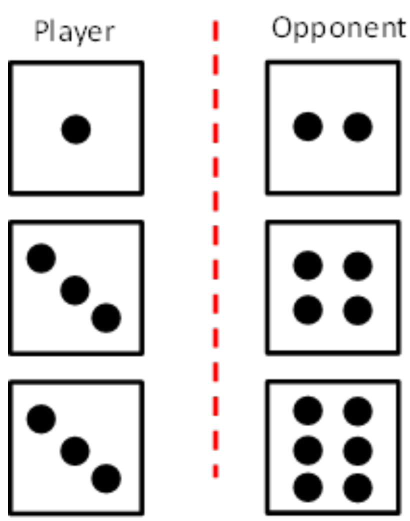
Out[32]= $\frac{12281}{46656}$

Out[33]= 0.263224

So we see that there is only a 26.3% chance of the bid being true so we should call the house’s bluff (effectively calling them a liar).

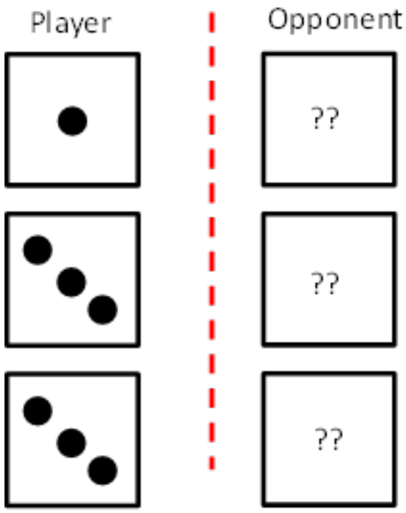
Let us now return to the game where we play against the opponent directly.

Assume that you have a hand of one, three, three
Assume your opponent has a hand of two, four, six



Suppose your opponent makes the same bid of 2 “fours” (meaning he claims that between the two of your dice sets, there are at least 2 “fours”).

Let us examine the situation from our perspective



From our perspective, we see that none of our die are a “four” (we hold a “one”, a “three”, and another “three”), so we see that in order for our opponents bid to be true, he must have at least two “fours”. In other words, in $n = 3$ independent trials (the opponent’s dice), there must be two successes or more.

In[34]:=
```
(*binomial distribution parameters from our perspective*)
n = 3;
p = 1 / 6;

(*Evaluate the distribution at the specified x values*)
PS = PDF[BinomialDistribution[n, p], {0, 1, 2, 3}]
```

Out[36]= $\left\{\frac{125}{216}, \frac{25}{72}, \frac{5}{72}, \frac{1}{216}\right\}$

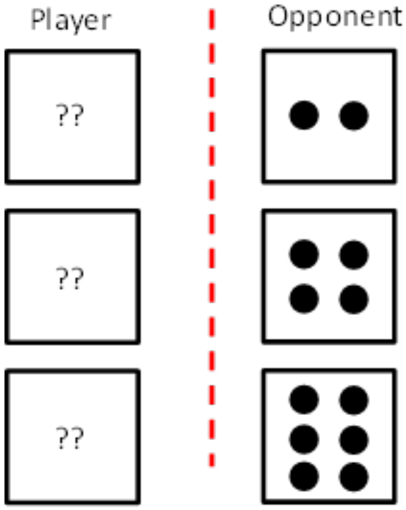
So from our perspective, we can calculate the probability that the bid is true.

In[37]:=
```
(*total cumulative sum that the bid is true*)
PS[[3]] + PS[[4]] // N
```

Out[37]= 0.0740741

So based on the information we have, we calculate that there is a 7.4% that the bid is true, so the action we should take based on this available information would be to call the bluff.

Let us examine the situation from the opponent’s perspective



From the opponent's perspective, he sees that he is holding onto 1 four, so if in our hand, we hold a 1 or more "fours" then the bid is true. In other words, in $n = 3$ independent trials (our dice), our opponent only needs one success or more. So from his perspective, the total probability that the bid is true is given by

In[38]:=
```
(*total cumulative sum that the bid is true*)
PS⟦2⟧ + PS⟦3⟧ + PS⟦4⟧ // N
```

Out[38]=
```
0.421296
```

So we see that the opponent is less likely to call the bluff as he calculates that there is a 42.1% chance of the bid being true.