Binomial Distribution

Binomial random variables arise in experiments with the following properties:

- The experiment consists of a fixed number, n, of Bernoulli trials each with probability p of success.
- The trials are identical and independent. The probability of success p, remains the same from trial to trial.
- The random variable X denotes ***the number of success obtained in the n trials***.

Let's take a look at a special case with $n = 3$:

The sample space is: $S = \{fff, sff, fsf, ffs, ssf, sfs, fss, sss\}$

X can take values from: {0,1,2,3}

$$f(0) = P[X = 0] = (1 - p)^3$$
$$f(1) = P[X = 1] = 3 \times (1 - p)^2 p$$
$$f(2) = P[X = 2] = 3 \times (1 - p) p^2$$
$$f(3) = P[X = 3] = p^3$$

i.e., $f(x) = \binom{3}{x} p^x (1 - p)^{3-x}, x = 0,1,2,3$

Can you makeup an experiment that generates a binomial random variable?

Density Function of Binomial Distribution

A random variable X has a binomial distribution with parameters $n\ and\ p$ if its density function is given by $f(x) = \binom{n}{x} p^x (1-p)^{n-x},\ 0 < p < 1,\ x = 0,1,2,3,\dots,\mathrm{n}$, where n is a positive integer.

Verify that this is a density function!

It is easy to verify that $f(x) \geq 0$.

Consider the *binomial theorem*: $(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^{n-k} b^k$

e.g., $n = 2: (a+b)^2 = a^2 + 2ab + b^2$

$n = 3: (a+b)^3 = a^3 + 3a^2b + 3ab^2 + b^3$

$n = 4: (a+b)^4 = a^4 + 4a^3b + 6a^2b^2 + 4ab^3 + b^4$

$\vdots$

$n = n: (a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$

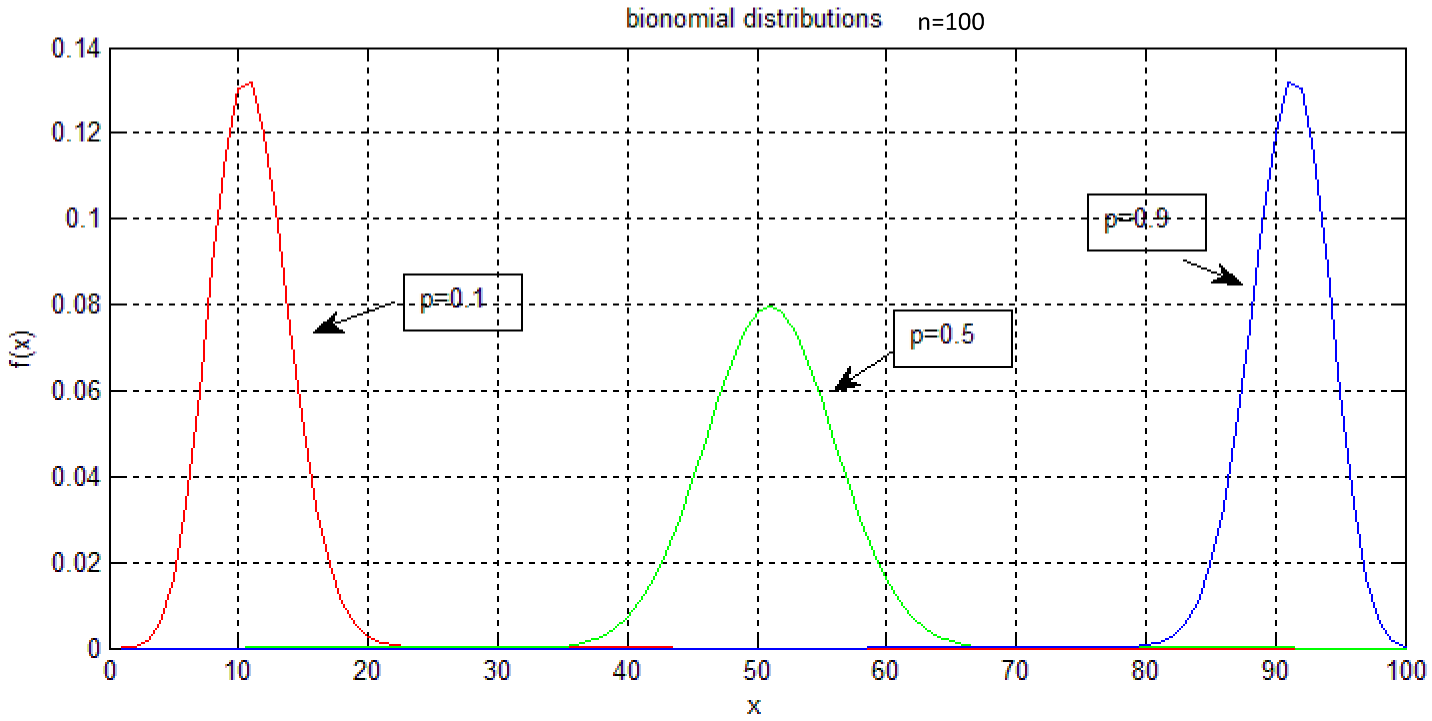
Then,

$$\sum_{x=0}^{n} f(x) = \sum_{x=0}^{n} \binom{n}{x} p^x (1-p)^{n-x} = [p + (1-p)]^n = 1$$

Binomial Distribution

In general, a random variable X has a binomial distribution with parameters $n\ and\ p$ if its density function is given by:

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}, x = 0,1,2,\ldots,n, \qquad 0 < p < 1, \qquad n\ is\ a\ positive\ integer$$



Guess the formula of $E[X]$

Binomial Distribution

In Class Assignment Problem #4:

The moment generating function of a binomial random variable X with parameter $p\ and\ n$ is

$$m_X(t) = (q + pe^t)^n, \qquad q = 1 - p$$

Find $E[X]\ and\ Var\ X$.

Solution:

$$\frac{d}{dt}m_X(t) = npe^t(q + pe^t)^{n-1} \Rightarrow \frac{d}{dt}m_X(t)_{t=0} = np(q + p)^n = np = E[X]$$

$$\frac{d^2}{dt^2}m_X(t) = npe^t(q + pe^t)^n + n(n-1)(pe^t)^2(q + pe^t)^{n-2} \Rightarrow \frac{d^2}{dt^2}m_X(t)_{t=0} = np + n(n-1)p^2 = E[X^2]$$

$$VarX = E[X^2] - (E[X])^2 = np - np^2 = np(1 - p)$$

Binomial Distribution Table

Table I gives the value of cumulative distribution

$$F(t) = P[X \le t] = \sum_{x=0}^{t} \binom{n}{x} p^x (1-p)^{n-x}$$

For selected values of n and p.

Example: Let X denote a binomial RV with $n = 9$ and $p = 0.5$. Find the probability $P[2 \le X \le 7]$ using Table I.

Solution:

$$P[2 \le X \le 7] = P[X \le 7] - P[X < 2]$$
$$= P[X \le 7] - P[X \le 1]$$
$$= F(7) - F(1)$$
$$= 0.9805 - 0.0195 = 0.9610$$

TABLE I
Cumulative binomial distribution

$$F_X(t) = P[X \le t] = \sum_{x \le t} \binom{n}{x} p^x (1-p)^{n-x}$$

		p										
n	t	0.1	0.2	0.25	0.3	0.4	0.5	0.6	0.7	0.75	0.8	0.9
5	0	0.5905	0.3277	0.2373	0.1681	0.0778	0.0312	0.0102	0.0024	0.0010	0.0003	0.0000
	1	0.9185	0.7373	0.6328	0.5282	0.3370	0.1875	0.0870	0.0308	0.0156	0.0067	0.0005
	2	0.9914	0.9421	0.8965	0.8369	0.6826	0.5000	0.3174	0.1631	0.1035	0.0579	0.0086
	3	0.9995	0.9933	0.9844	0.9692	0.9130	0.8125	0.6630	0.4718	0.3672	0.2627	0.0815
	4	1.0000	0.9997	0.9990	0.9976	0.9898	0.9688	0.9222	0.8319	0.7627	0.6723	0.4095
	5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
6	0	0.5314	0.2621	0.1780	0.1176	0.0467	0.0156	0.0041	0.0007	0.0002	0.0001	0.0000
	1	0.8857	0.6554	0.5339	0.4202	0.2333	0.1094	0.0410	0.0109	0.0046	0.0016	0.0001
	2	0.9841	0.9011	0.8306	0.7443	0.5443	0.3437	0.1792	0.0705	0.0376	0.0170	0.0013
	3	0.9987	0.9830	0.9624	0.9295	0.8208	0.6562	0.4557	0.2557	0.1694	0.0989	0.0159
	4	0.9999	0.9984	0.9954	0.9891	0.9590	0.8906	0.7667	0.5798	0.4661	0.3446	0.1143
	5	1.0000	0.9999	0.9998	0.9993	0.9959	0.9844	0.9533	0.8824	0.8220	0.7379	0.4686
	6	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
7	0	0.4783	0.2097	0.1335	0.0824	0.0280	0.0078	0.0016	0.0002	0.0001	0.0000	0.0000
	1	0.8503	0.5767	0.4449	0.3294	0.1586	0.0625	0.0188	0.0038	0.0013	0.0004	0.0000
	2	0.9743	0.8520	0.7564	0.6471	0.4199	0.2266	0.0963	0.0288	0.0129	0.0047	0.0002
	3	0.9973	0.9667	0.9294	0.8740	0.7102	0.5000	0.2898	0.1260	0.0706	0.0333	0.0027
	4	0.9998	0.9953	0.9871	0.9712	0.9037	0.7734	0.5801	0.3529	0.2436	0.1480	0.0257
	5	1.0000	0.9996	0.9987	0.9962	0.9812	0.9375	0.8414	0.6706	0.5551	0.4233	0.1497
	6	1.0000	1.0000	0.9999	0.9998	0.9984	0.9922	0.9720	0.9176	0.8665	0.7903	0.5217
	7	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
8	0	0.4305	0.1678	0.1001	0.0576	0.0168	0.0039	0.0007	0.0001	0.0000	0.0000	0.0000
	1	0.8131	0.5033	0.3671	0.2553	0.1064	0.0352	0.0085	0.0013	0.0004	0.0001	0.0000
	2	0.9619	0.7969	0.6785	0.5518	0.3154	0.1445	0.0498	0.0113	0.0042	0.0012	0.0000
	3	0.9950	0.9437	0.8862	0.8059	0.5941	0.3633	0.1737	0.0580	0.0273	0.0104	0.0004
	4	0.9996	0.9896	0.9727	0.9420	0.8263	0.6367	0.4059	0.1941	0.1138	0.0563	0.0050
	5	1.0000	0.9988	0.9958	0.9887	0.9502	0.8555	0.6846	0.4482	0.3215	0.2031	0.0381
	6	1.0000	0.9999	0.9996	0.9987	0.9915	0.9648	0.8936	0.7447	0.6329	0.4967	0.1869
	7	1.0000	1.0000	1.0000	0.9999	0.9993	0.9961	0.9832	0.9424	0.8999	0.8322	0.5695
	8	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
9	0	0.3874	0.1342	0.0751	0.0404	0.0101	0.0020	0.0003	0.0000	0.0000	0.0000	0.0000
	1	0.7748	0.4362	0.3003	0.1960	0.0705	0.0195	0.0038	0.0004	0.0001	0.0000	0.0000
	2	0.9470	0.7382	0.6007	0.4628	0.2318	0.0898	0.0250	0.0043	0.0013	0.0003	0.0000
	3	0.9917	0.9144	0.8343	0.7297	0.4826	0.2539	0.0994	0.0253	0.0100	0.0031	0.0001
	4	0.9991	0.9804	0.9511	0.9012	0.7334	0.5000	0.2666	0.0988	0.0489	0.0196	0.0009
	5	0.9999	0.9969	0.9900	0.9747	0.9006	0.7461	0.5174	0.2703	0.1657	0.0856	0.0083
	6	1.0000	0.9997	0.9987	0.9957	0.9750	0.9102	0.7682	0.5372	0.3993	0.2618	0.0530
	7	1.0000	1.0000	0.9999	0.9996	0.9962	0.9805	0.9295	0.8040	0.6997	0.5638	0.2252
	8	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.9980	0.9899	0.9596	0.9249	0.8658	0.6126

Negative Binomial Distribution

Negative binomial random variables arise in experiments with the following properties:

- The experiment consists of a series of independent and identical Bernoulli trials each with probability p of success.
- The trials are observed until exactly r successes are obtained, where r is fixed.
- The random variable X denotes ***the number of trials needed to obtain the r successes***.

Let's consider the case of $r = 3$. X takes values from {3,4,5,…}
Typical outcomes can be: $\{sss, ssfs, ssffs, sfffss, ffffsss, \ldots\}$

Each outcome must end with a successful trial;
The remaining $x - 1$ trials must result in exactly two successes and $x - 3$ failures in some order;
Different outcomes can yield identical values for X. The number of outcomes that result in a given value x is $\binom{x-1}{2}$.
Therefore, $f(x) = P[X = x] = \binom{x-1}{2}(1-p)^{x-3}p^3, x = 3,4,5, \ldots$

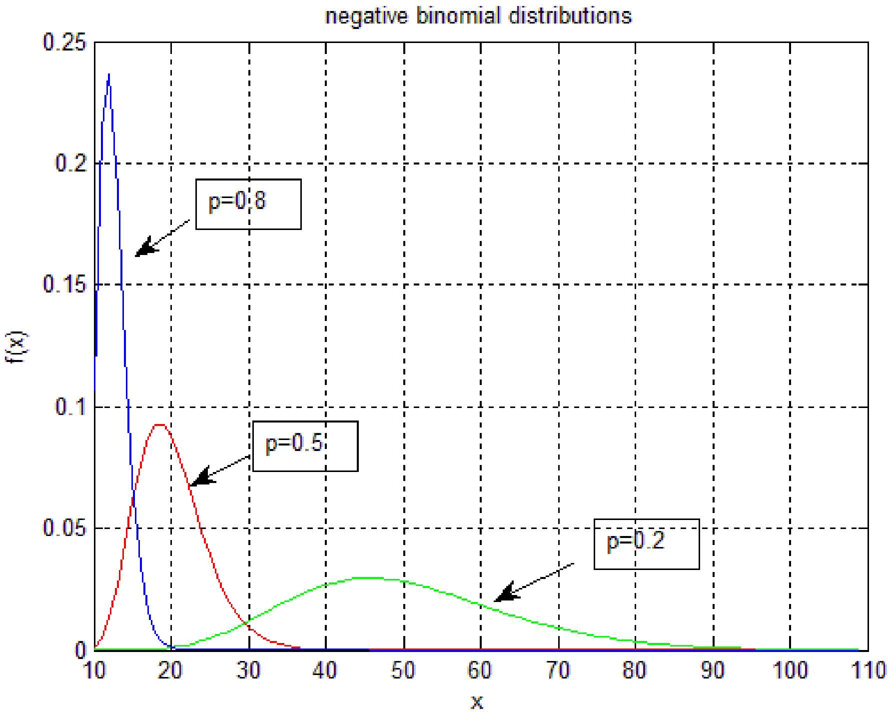
Eg, when $x = 4$, there exist three ,i.e.,$\binom{3}{2}$, different outcomes: $\{fsss, sfss, ssfs\}$

when $x = 5$, there exist six, i.e., $\binom{4}{2}$, different outcomes:$\{ffsss, fsfss, fssfs, sffss, sfsfs, ssffs\}$

Negative Binomial Distribution

A random variable is said to have negative binomial distribution with parameter $p\ and\ r$ if its density function is given by:

$$f(x) = \binom{x-1}{r-1}(1-p)^{x-r}p^r, \qquad r = 1,2,3,\dots;\ x = r, r+1, r+2, \dots$$



What is the relationship between geometric and negative binomial distribution?

What is the r value in the figure?
Can you guess the expected value of X?

Negative Binomial Distribution

Practice problem:

The moment generating function of a negative binomial random variable X with parameters $p\ and\ r$ is

$$m_X(t) = \left(\frac{pe^t}{1-qe^t}\right)^r, \qquad q = 1-p$$

Find $E[X]\ \ and\ VarX$

Solution:

$$E[X] = \frac{r}{p}, VarX = \frac{r(1-p)}{p^2}$$

$$\frac{d}{dt}m_X(t) = r\left(\frac{pe^t}{1-qe^t}\right)^{r-1}\frac{pe^t}{(1-qe^t)^2} \Longrightarrow \frac{d}{dt}m_X(t)\Big|_{t=0} = \frac{r}{p} = E[X]$$

$$\frac{d^2}{dt^2}m_X(t) = r(r-1)\left(\frac{pe^t}{1-qe^t}\right)^{r-2}\left[\frac{pe^t}{(1-qe^t)^2}\right]^2 + r\left(\frac{pe^t}{1-qe^t}\right)^{r-1}\frac{pe^t(1+qe^t)}{(1-qe^t)^3}$$

$$\Longrightarrow \frac{d^2}{dt^2}m_X(t)\Big|_{t=0} = \frac{r(1+q)+r^2-r}{p^2} = E[X^2]$$

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

Poisson Distribution

Poisson random variables arise in experiments with the following properties:

- The experiment consists of counting events occur in an interval $(t, t+\tau)$.
- The interval can be of ***time or space***
- The occurrences are independent.
- The events occur at a constant average rate, λ, also known as intensity.
- The number of events occur in the interval follows the Poisson distribution with the following density function:

$$f(x) = P[X = x] = \frac{e^{-\lambda\tau}(\lambda\tau)^x}{x!} = \frac{e^{-k}(k)^x}{x!}, \qquad k = \lambda\tau, x = 0,1,2,3,\ldots$$

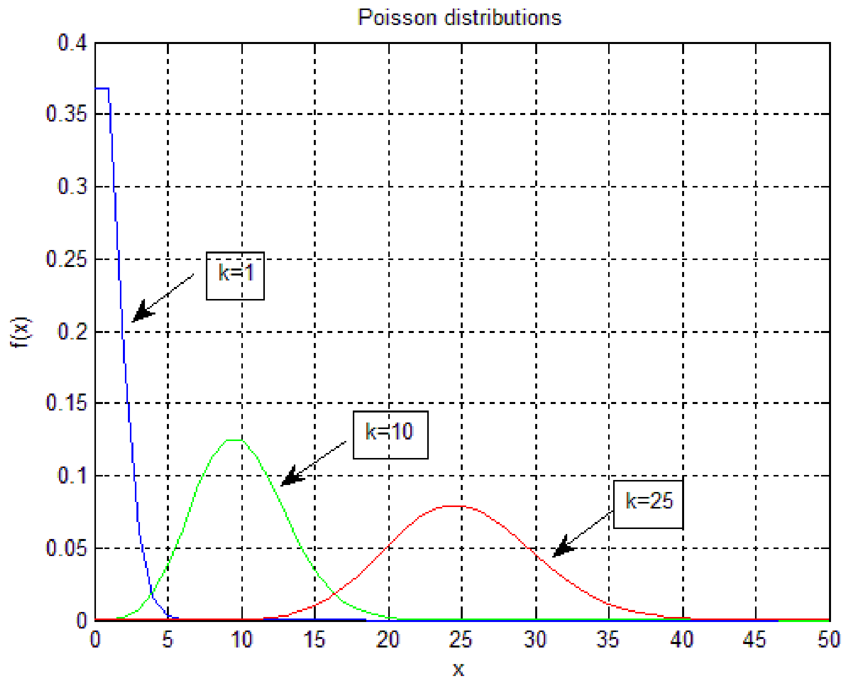
$$f(0) = \qquad P[X = 0] = e^{-k}$$

Prove that $f(x)$ is a density function.

Examples of Poisson RVs:

- # of mails received in a day
- # of phone call received in an hour
- # of customers come to an ATM in an hour
- # of vehicles passing a check point in a minute

⋮



How are E[X] and VarX related to k?

Poisson Distribution

Prove that

$$f(x) = P[X = x] = \frac{e^{-\lambda\tau}(\lambda\tau)^x}{x!} = \frac{e^{-k}(k)^x}{x!}, \qquad k = \lambda\tau, x = 0,1,2,3,$$

is a density function.

Proof:

(1) It is easy to verify that $f(x) \geq 0$, $x = 0,1,2,3, \ldots$

(2) For $x = 0,1,2,3, \ldots$, we have,

$$\sum_{x=0}^{\infty} f(x) = \sum_{x=0}^{\infty} \frac{e^{-k}(k)^x}{x!} = e^{-k} \sum_{x=0}^{\infty} \frac{(k)^x}{x!} = e^{-k}e^{k} = 1$$

Poisson Distribution

Practice Example:

The moment generating function (m.g.f) $m_X(t)$ of a Poisson random variable X with parameter $k = \lambda\tau$ is $m_X(t) = e^{k(e^t - 1)}$. Use the m.g.f to calculate $E[X]\ and\ VarX$.

Solution:

$$\frac{d}{dt}m_x(t) = ke^t e^{(ke^t - k)} = ke^{(ke^t + t - k)} \Rightarrow \frac{d}{dt}m_x(t)_{t=0} = k = E[X]$$

$$\frac{d^2}{dt^2}m_x(t) = (ke^t + 1)ke^{(ke^t + t - k)} \Rightarrow \frac{d^2}{dt^2}m_x(t)_{t=0} = k(k+1) = E[X^2]$$

$$VarX = E[X^2] - (E[X])^2 = k(k+1) - k^2 = k$$

$$E[X] = k, E[X^2] = k(k+1), VarX = k$$

Poisson Distribution

Example: The white blood cell count of a healthy individual follows a Poisson distribution and can average as low as 6000 per cubic millimeter of blood. To detect a white-cell deficiency, a 0.001 cubic millimeter drop of blood is taken and the number of white cells X is found.
(1) How many white blood cells are expected in a drop of blood from a healthy individual?
(2) If at most two are found, is this evidence of a white cell deficiency?

Solution:
This experiment can be viewed as involving a Poisson process. The discrete event of interest is the occurrence of a white cell. The continuous space interval is a drop of blood.
Then,

$$\lambda = 6000; \tau = 0.001; k = \lambda\tau = 6; E[X] = k = 6$$

i.e., in a healthy individual, we would expect, on average, to see 6 white cells in a drop of blood

We want to find $P[X \leq 2]$:

$$P[X \leq 2] = \sum_{x=0}^{2} f(x) = \sum_{x=0}^{2} \frac{e^{-6}6^x}{x!} = \frac{e^{-6}6^0}{0!} + \frac{e^{-6}6^1}{1!} + \frac{e^{-6}6^2}{2!} \approx 0.062$$

Since this is a small probability, it is reasonable to consider that this individual has a white cell deficiency!!!

Discrete Distributions

Distributions	Density function	Moment generating function
Geometric	$(1-p)^{x-1}p$	$\frac{pe^t}{1-qe^t}$
Binomial	$\binom{n}{x}p^x(1-p)^{n-x}$	$(q+pe^t)^n$
Negative Binomial	$\binom{x-1}{r-1}(1-p)^{x-r}p^r$	$\left(\frac{pe^t}{1-qe^t}\right)^r$
Poisson	$\frac{e^{-k}(k)^x}{x!}, k=\lambda\tau$	$e^{k(e^t-1)}$