Week 4: Variance, moments, geometric, binomial and Poisson distributions

Armenak Petrosyan

Last time

- We defined the **mathematical expectation** (also called **mean value** or **expected value**) of the random variable X as

$$\mu = E[X] = \sum_{x \in \mathrm{Range}(X)} x f(x).$$

- It generalizes the arithmetic mean to random variables.
- If $Y = u(X)$, then

$$E[Y] = \sum_{x \in \mathrm{Range}(X)} u(x) f(x).$$

Hypergeometric distribution

Theorem

Let X be a hypergeometric distribution with parameters (N, K, n). Then its mean is equal to

$$\mu = \frac{nK}{N}.$$

Intuitively: red balls are the $\frac{K}{N}$ part of all balls. We are doing n selections so expect to get on average $n\frac{K}{N}$ red balls.

Proof

- From definition and the formula for the range and pmf of hypergeometric distribution

$$\begin{aligned} E[X] &= \sum_{x \in \mathrm{Range}(X)} x f(x) \\ &= \sum_{x=\max\{0, n-(N-K)\}}^{\min\{n,K\}} x \frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}} \\ &= \sum_{x=\max\{1, n-(N-K)\}}^{\min\{n,K\}} x \frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}}. \end{aligned}$$

Proof (cont.)

- You can check that

$$x\frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}}=\frac{nK}{N}\frac{\binom{K-1}{x-1}\binom{(N-1)-(K-1)}{(n-1)-(x-1)}}{\binom{N-1}{n-1}}.$$

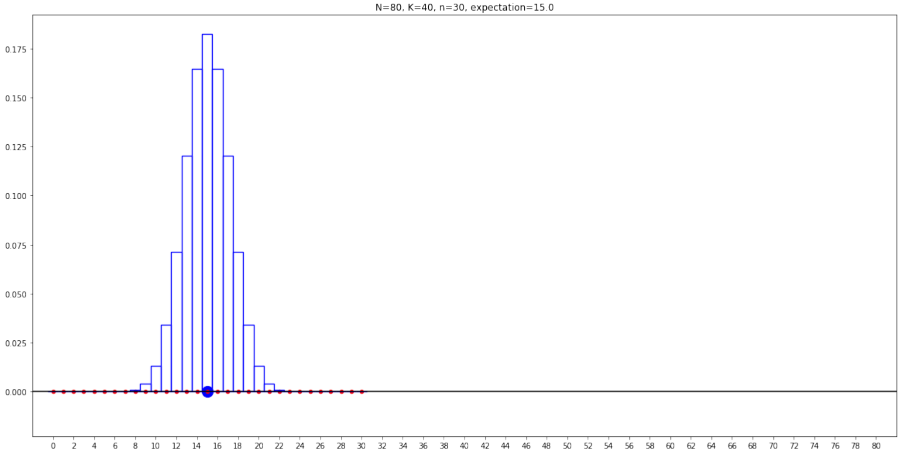
- Thus

$$\begin{aligned}E[X]&=\sum_{x=\max\{1,n-(N-K)\}}^{\min\{n,K\}}\frac{nK}{N}\frac{\binom{K-1}{x-1}\binom{(N-1)-(K-1)}{(n-1)-(x-1)}}{\binom{N-1}{n-1}}\\&=\frac{nK}{N}\sum_{y=\frac{nK}{N}\max\{1,n-1-((N-1)-(K-1))\}}^{\min\{n-1,K-1\}}\frac{\binom{K-1}{y}\binom{(N-1)-(K-1)}{(n-1)-y}}{\binom{N-1}{n-1}}\end{aligned}$$

  after changing $x-1$ with y.
- Notice that

$$\sum_{y=\frac{nK}{N}\max\{1,n-1-((N-1)-(K-1))\}}^{\min\{n-1,K-1\}}\frac{\binom{K-1}{y}\binom{(N-1)-(K-1)}{(n-1)-y}}{\binom{N-1}{n-1}}$$

  is equal to 1 as the sum of all pmf values of a hypergemetric distribution with parameters $(N-1,K-1,n-1)$.


N=80, K=40, n=30, expectation=15.0
0.175
0.150
0.125
0.100
0.075
0.050
0.025
0.000
0 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30 32 34 36 38 40 42 44 46 48 50 52 54 56 58 60 62 64 66 68 70 72 74 76 78 80

Observation

- Put $u(x) = (x-b)^2$: measures distance between x and b.
- It is the square of $|x-b|$ and easier to work with.
- $E[(X-b)^2] =$ "average" distance between X and b.
- Notice
$$g(b) = E[(X-b)^2] = E[X^2 - 2bX + b^2] = E[X^2] - 2bE[X] + b^2.$$
- Let us compute the minimum of the function $g(b)$:
$$\frac{\partial g}{\partial b}(b) = 2b - 2E[X] = 0$$
hence the minimum is at $b = E[X]$.

Intuitively: expected value is the "center" of the histogram where it concentrates (the point all of the values of the random variable are simultaneously close to). See the hyper-geometric example above.

Tuesday class

Geometric distribution

- A random trial has a probability of success equal to p and probability of failure $q = 1 - p$.
- Consider the following experiment: we are doing consecutive random trials until we reach a success.
- The set of outcomes has the form $s = \overbrace{FF \cdots F}^{x-1} S$ where number of F's can be any number $x = 1, 2, \ldots$
- Then $P(s) = q^{x-1}p$.
- Let $X(s)$ denote the number of trials it took to reach success:

$$X(\overbrace{FF \cdots F}^{x-1} S) = x.$$

- Note that

$$f(x) = q^{x-1}p.$$

Definition (Geometric distribution)

The pmf of the random variable X is called **Geometric distribution**.

Theorem

If X has the Geometric distribution, then

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

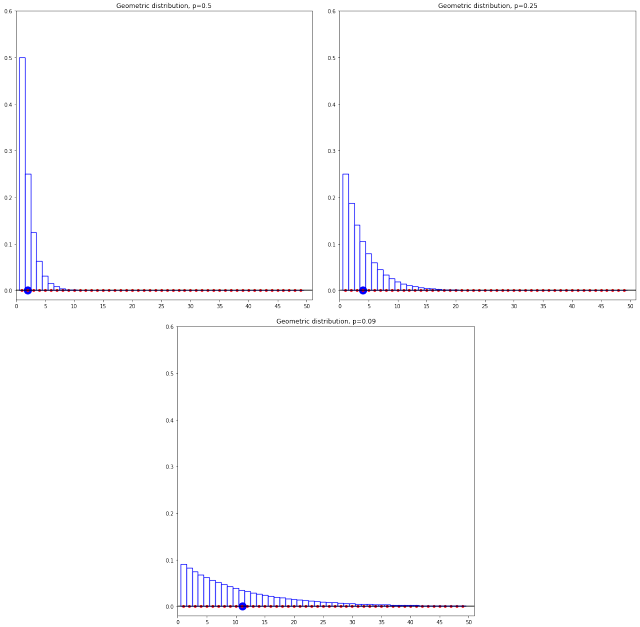
Intuitively, if the success rate is $p = \frac{1}{10}$, then in average it takes 10 trials to reach success.

Proof.

- $E[X] = \sum_{x=1}^{\infty} x f(x) = \sum_{x=1}^{\infty} x q^{x-1} p.$
-

$$\begin{aligned} pE[X] &= (1-q)E[X] = E[X] - qE[X] \\ &= \sum_{x=1}^{\infty} x q^{x-1} p - q \sum_{x=1}^{\infty} x q^{x-1} p = \\ &= \sum_{x=1}^{\infty} x q^{x-1} p - \sum_{x=1}^{\infty} x q^{x} p = \sum_{x=1}^{\infty} x q^{x-1} p - \sum_{x=1}^{\infty} (x-1) q^{x-1} p \\ &= \sum_{x=1}^{\infty} q^{x-1} p = \frac{1}{1-q} p = 1. \end{aligned}$$

$\square$


Geometric distribution, p=0.5
Geometric distribution, p=0.25
Geometric distribution, p=0.09

Variance and standard deviation

- We observed that, in certain sense, most of the values of X are concentrated around the mean $\mu = E[X]$.
- To measure how much the values of X deviate from the mean, we use the quantity

$$\mathrm{Var}(X) = E[(X - E[X])^2].$$

Definition (Variance)

$\mathrm{Var}(X)$ is called the **variance** of the random variable X.

Definition (Standard deviation)

The square root of variance is called the **standard deviation** of X:

$$\sigma = \sqrt{E[(X - E[X])^2]}.$$

- The larger the variance is, the more "spread out" the histogram of the random variable is.

Theorem

Variance can be computed also as

$$\sigma^2 = E[X^2] - E[X]^2.$$

Proof.

$$\begin{aligned}\sigma^2 &= E[(X - E[X])^2] = E[X^2 - 2E[X]X + E[X]^2] \\ &= E[X^2] - 2E[X]E[X] + E[X]^2 \\ &= E[X^2] - E[X]^2.\end{aligned}$$

□

Exercise 1

Problem (2.3-8 in the textbook)

Let X equal the larger outcome when a pair of fair four-sided dice is rolled. The pmf of X is

$$f(x) = \frac{2x-1}{4^2}, \quad x = 1, 2, 3, 4.$$

Find the mean, variance, and standard deviation of X.

Proof.

- $E[X] = 1f(1) + 2f(2) + 3f(3) + 4f(4) = \frac{1}{16} + \frac{6}{16} + \frac{15}{16} + \frac{28}{16} = \frac{50}{16} = \frac{25}{8}$.
- $E[X^2] = 1f(1) + 2^2 f(2) + 3^2 f(3) + 4^2 f(4) = \frac{1}{16} + \frac{12}{16} + \frac{45}{16} + \frac{112}{16} = \frac{170}{16} = \frac{85}{8}$.
- $\mathrm{Var}(X) = E[X^2] - E[X]^2 = \frac{85}{8} - \frac{625}{64} = \frac{680-625}{64} = \frac{55}{64}$.
- $\sigma = \sqrt{\frac{55}{64}}$.

$\square$

Theorem

Suppose X is a random variable with mean μ_X and standard deviation σ_X. Let $Y = aX + b$ where a and b are any two numbers. Then mean and the standard deviation of Y are

$$\mu_Y = a\mu_X + b, \quad \sigma_Y = |a|\sigma_X.$$

Proof.

- $\mu_Y = E[aX + b] = aE[X] + b = a\mu_X + b.$
- Notice that

$$\begin{aligned}
\mathrm{Var}(Y) &= E[(Y - E[Y])^2] \\
&= E[(aX + b - aE[X] - b)^2] \\
&= E[(aX - aE[X])^2] \\
&= a^2\mathrm{Var}(X).
\end{aligned}$$

- Taking square roots, we have $\sigma_Y = |a|\sigma_X$.

$\square$

Example

- Let X be a random variable with $\mathrm{Range}(X) = \{-2, -1, 0, 1, 2\}$ and

$$f_X(-2) = 0.4, \quad f_X(-1) = 0.25, \quad f_X(0) = 0.15, \quad f_X(1) = 0.1, \quad f_X(2) = 0.1.$$

- Take the random variable $Y = 2X + 1$.
- $\mathrm{Range}(Y) = \{-3, -1, 1, 3, 5\}$ and

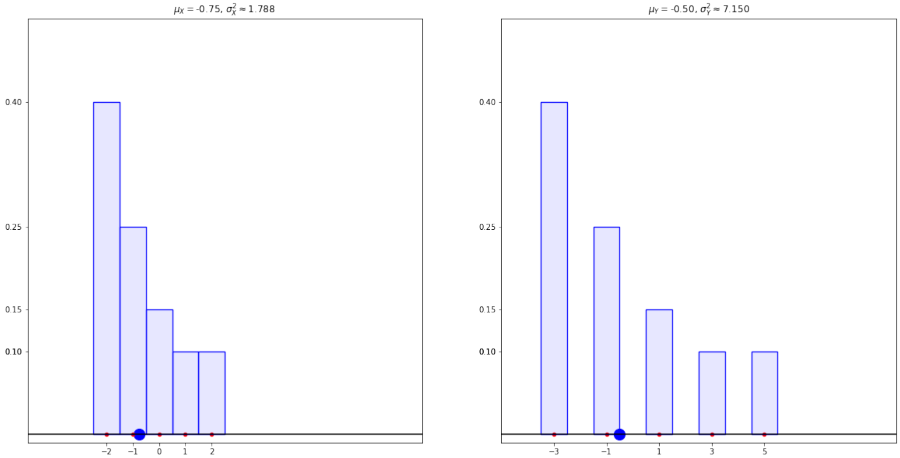
$$f_Y(-3) = 0.4, \quad f_Y(-1) = 0.25, \quad f_Y(1) = 0.15, \quad f_Y(3) = 0.1, \quad f_Y(5) = 0.1.$$

- It can be checked that

$$\mu_X = -0.75, \quad \sigma_X \approx 1.337.$$

- And, thus

$$\mu_Y = -0.5, \quad \sigma_Y \approx 2.674.$$



$\mu_X = -0.75$, $\sigma_X^2 \approx 1.788$
0.40
0.25
0.15
0.10
−2
−1
0
1
2
$\mu_Y = -0.50$, $\sigma_Y^2 \approx 7.150$
0.40
0.25
0.15
0.10
−3
−1
1
3
5

Definition

Let $r \in \mathbb{N}$. If the following mathematical expectations exist, they are called

- $E[X^r]$ – r**-th moment of** X.
- $E[(X-b)^r]$ – r**-th moment of** X **around** b.
- $E[X-E[X])^r]$ – **central** r**-th moment of** X.
- $\frac{E[(X-E[X])^r]}{\sigma^r}$ – **standardized** r**-th moment of** X.
- $E[X(X-1)\cdots(X-r+1)]$ – **factorial** r**-th moment of** X.

Theorem

$$\mathrm{Var}(X) = E[X(X-1)] + E[X] - E[X]^2.$$

Proof.

- $E[X(X-1)] = E[X^2 - X] = E[X^2] - E[X]$.
- And so $E[X(X-1)] + E[X] - E[X]^2 = E[X^2] - E[X]^2 = \mathrm{Var}(X)$.

$\square$

Exercise 2

For certain distributions, factorial moments are easier to compute.

Problem (2.3-10 in the textbook)

Let X be a random variable that has hypergeometric distribution with parameters (N, K, n). Then

$$E[X(X-1)] = \frac{n(n-1)K(K-1)}{N(N-1)}.$$

Solution

Similar argument as in the computation of $E[X]$ for the hypergeometric distribution (left as homework) .

Theorem

Let X be a hypergeometric distribution with parameters (N, K, n). Then

$$\mathrm{Var}(X) = n\frac{K}{N}\left(1 - \frac{K}{N}\right)\frac{N-n}{N-1}.$$

Proof.

$$\begin{aligned}
\mathrm{Var}(X) &= E[X(X-1)] + E[X] - E[X]^2 \\
&= \frac{n(n-1)K(K-1)}{N(N-1)} + n\frac{K}{N} - n^2\frac{K^2}{N^2} \\
&= n\frac{K}{N}\left(1 - \frac{K}{N}\right)\frac{N-n}{N-1}.
\end{aligned}$$

$\square$

Why do we care about moments in practice?

- They provide easy to compute numerical summaries of the distribution.
- Instead of looking at the whole histogram, you extract a few numbers.
- For example, for the hyper-geometric distribution, the number of parameters is 3 (they are the N, K, n) so knowing only the mean and variance does not provide enough information to find the parameters, if they are unknown. You need at least 3 moments.
- The approach of identifying parameters in the distribution using moments is called **the method of moments**.

Exercise 3

Problem (Mark and recapture method)

The game fish manager is trying to find the number of fish in a pond. To do that he capture 50 fish marks them, and releases back into the pond. Then every day for two months he randomly catches five different fish, records the number of marked fish in that day's sample, and releases them back into the pond. His records are displayed below

$$D = \{1,0,1,1,3,0,2,1,0,0,1,0,0,0,0,0,0,2,0,0,0,1,1,1,1,0,1,1,0,1,\\ 2,1,1,1,0,1,0,0,0,0,0,0,1,0,0,0,0,0,0,2,0,0,1,1,0,0,1,0,1,0\}$$

What is the number of fish in the pond.

Solution

- Let X be the number of marked fish in the sample.
- X is a hypergeometric distribution with parameters $(N, K = 50, n = 5)$.
- The empirical mean is equal to $\bar{\mu} = \dfrac{\sum\limits_{x \in D} x}{\#\text{ of elements in } D} \approx 0.53.$.
- Using the formula for the mean of the Hypergeometric distribution

$$\mu = \frac{nK}{N} = \frac{250}{N} \approx 0.53 \Rightarrow N \approx 471 \text{ (compare to } 450).$$

.

Definition (Moment generating function)

Let X be a discrete random variable with pmf $f(x)$. If the mathematical expectation

$$E[e^{tX}] = \sum_{x \in \mathrm{Range}(X)} e^{tx} f(x)$$

exists for every t in some interval $(-h, h)$ then the function

$$M(t) = E[e^{tX}]$$

is called the **moment generating function** of X **(mgf)**.

FACT: if two discrete random variables have the same mgf then they have the same pmf.

Theorem

If the moment generating function of X exists then

$$M^{(r)}(0) = E[X^r].$$

Proof.

- From the assumption that the mathematical expectation exists in $(-h, h)$, we can do

$$\frac{d^r}{dt^r} M(0) = \sum_{x \in \mathrm{Range}(X)} \frac{d^r}{dt^r} e^{tx}|_{t=0} f(x)$$

- $\frac{d^r}{dt^r} e^{tx}|_{t=0} = x^r$ (check from Taylor series of $e^{tx} = \sum_{i=0}^{\infty} \frac{(tx)^i}{i!}$).
- Thus

$$\frac{d^r}{dt^r} M(0) = \sum_{x \in \mathrm{Range}(X)} x^r f(x) = E[X^r].$$

$\square$

Exercise 4

Problem (2.5-7 from the textbook)

If $E[X^r] = 5^r$, $r = 1, 2, \ldots$, find the moment generating function $M(t)$ of X and the pmf of X.

Solution

- *We saw that $E[X^r] = M^r(0)$, therefore from Taylor expansion*

$$M(t) = \sum_{r=0}^{\infty} \frac{M^r(0)}{r!} x^r = M(0) + \sum_{r=1}^{\infty} \frac{5^r}{r!} x^r = 1 + \sum_{r=1}^{\infty} \frac{5^r}{r!} x^r = e^{5x}.$$

- *We used the fact that $M(0) = \sum_{x \in \mathrm{Range}(X)} f(x) = 1$.*
- *Because pmf is uniquely determined by the mgf, and $M(t)$ is the mgf of the following pmf*

$$f(x) = \begin{cases} 1 & x = 5 \\ 0 & \textit{otherwise} \end{cases} \quad \textit{for } x = 1, 2, \ldots$$

we conclude that this must be the pmf of X.

Example

- Let X be a geometric distribution. with parameter p and let $q = 1 - p$.
- $\mathrm{Range}(f) = \{1, 2, \ldots\}$ and $f(x) = q^{x-1}p$.
- Then

$$M(t) = \sum_{x=1}^{\infty} e^{xt} q^{x-1} p = \frac{p}{q} \sum_{x=1}^{\infty} (e^t q)^x = \frac{p}{q} \sum_{x=1}^{\infty} (e^t q)^x = \frac{p}{q} \frac{e^t q}{1 - e^t q} = \frac{e^t p}{1 - e^t q}.$$

-

$$M'(t) = \frac{(1 - qe^t)pe^t - (-qe^t)pe^t}{(1 - e^t q)^2} = \frac{pe^t}{(1 - e^t q)^2}.$$

- Similar calculations yield

$$M''(t) = \frac{pe^t(1 + qe^t)}{(1 - e^t q)^3}.$$

- Therefore

$$E[X] = M'(0) = \frac{p}{(1 - q)^2} = \frac{1}{p}$$

$$Var(X) = M''(0) - (M'(0))^2 = \frac{p(1 + q)}{(1 - q)^3} - \frac{1}{p} = \frac{p(1 + q)}{p^3} - \frac{1}{p} = \frac{q}{p^2}.$$

Thursday class

cdf of the geometric distribution

- Let X have the geometric distribution with parameter p and let $q = 1 - p$.
- Notice that
$$P[X > k] = \sum_{x=k+1}^{\infty} q^{x-1} p = pq^k \sum_{x=0}^{\infty} q^x = \frac{pq^k}{1-q} = q^k.$$
- Therefore, for any $k = 1, 2, \ldots,$
$$F(k) = P[X \leq k] = 1 - P[X > k] = 1 - q^k.$$

Definition (Bernoulli experiment)

In **Bernoulli experiment** the outcome is either of Type I (success) with probability p or Type II (fail) with probability $q = 1 - p$.

Bernoulli trials are the outcomes of several independent Bernoulli experiments with the same success probabilities.

Example

- Consider the coin flip experiment with a biased coin that lands head with probability p and tail with probability $q = 1 - p$.
- A single coin flip will be a Bernoulli experiment.
- n independent coin flips will be a Bernoulli trial.

Definition (Binomial distribution)

Let X denote the number of successes in n Bernoulli trials with success probability p. Then we say that X has the **binomial distribution** with parameters (n, p).

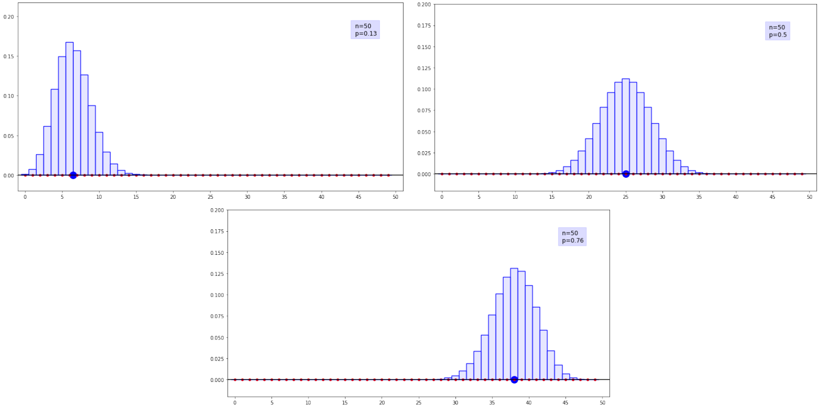
If X has binomial distribution with parameters (n, p) then

1. $\mathrm{Range}(X) = \{0, 1, \ldots, n\}$.
2. For any $x = 0, 1, 2, \ldots, n$,

$$f(x) = \binom{n}{x} p^x q^{n-x}$$

$\binom{n}{x}$ ways there can be x successes: the rest $n - x$ are failures.

Binomial histogram



mgf of binomial distribution

From the definition of mgf:

$$
\begin{aligned}
M(t) = E[e^{tX}] &= \sum_{x=0}^{n} e^{tx} \binom{n}{x} p^x q^{n-x} \\
&= \sum_{x=0}^{n} \binom{n}{x} (pe^t)^x q^{n-x} \\
&\text{(using the the binomial theorem)} \\
&= [q + pe^t]^n \\
&= [1 - p + pe^t]^n.
\end{aligned}
$$

Mean and variance of the binomial distribution

- $$M'(t) = npe^t[1 - p + pe^t]^{n-1}$$

  and so

  $$E[X] = M'(0) = np.$$

  Intuition: if probability of success in one trial is p, the probability of success in n trials will be n times more.

- $$M''(t) = npe^t[1 - p + pe^t]^{n-1} + n(n-1)p^2e^{2t}[1 - p + pe^t]^{n-2}$$

  and so

  $$E[X^2] = M''(0) = np + n(n-1)p^2.$$

- Finally,

$$\begin{aligned} \mathrm{Var}(X) &= E[X^2] - E[X]^2 \\ &= np + n(n-1)p^2 - n^2p^2 \\ &= np - np^2 = np(1-p) = npq. \end{aligned}$$

Exercise 5

Problem (2.4-13 in the textbook)

It is claimed that for a particular lottery, 1/10 of the 50 million tickets will win a prize. What is the probability of winning at least one prize if you purchase

(a) *10 tickets*

(b) *15 tickets.*

Solution

1. - *Let* X *be the number of winning tickets among* 10 *bought. Then* X *is a binomial distribution with parameters* $(n = 10, p = 0.1)$.
   - *We want to find*

$$P(X \geq 1) = 1 - f(0) = 1 - \binom{10}{0} p^0 q^{10-0} = 1 - 0.9^{10} \approx 0.65.$$

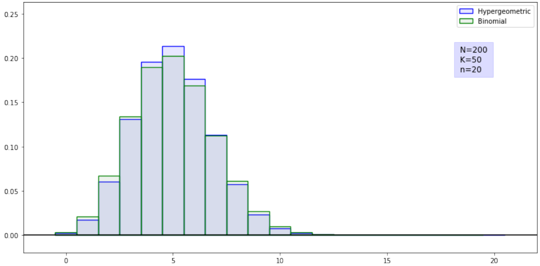
2. - *Similarly, for* $n = 15$ *and* $p = 0.1$,

$$P(X \geq 1) = 1 - f(0) = 1 - \binom{15}{0} p^0 q^{15-0} = 1 - 0.9^{15} \approx 0.79.$$

Binomial vs hypergeometric

Suppose there are K red balls and $N-K$ blue balls in an urn and we are sampling n of them.

- If the sampling is done without replacement then the number of red balls in the sample is a random variable that has hypergeometric distribution with parameters (N, K, n).
- If the sampling is done with replacement then the number of red balls in the sample is a random variable that has binomial distribution with parameters (n, p) where $p = \frac{K}{N}$.
- When $N >> n$, they are very close to each other.



cdf of the binomial distribution

The cdf of the binomial distribution can be computed as

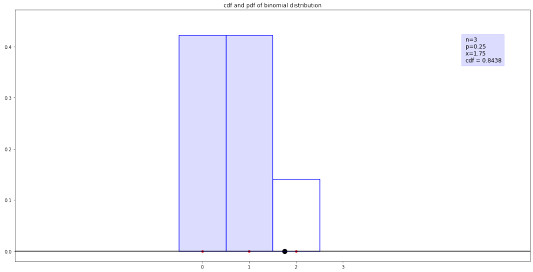
$$F(x) = E[X \leq x] = \sum_{y=0}^{\lfloor x \rfloor} \binom{n}{x} p^x q^{n-x} \text{ for } x \in [0, n].$$

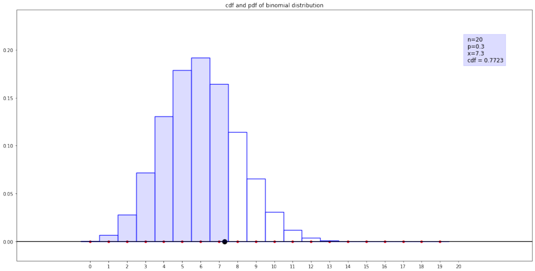
- $F(x)$ does not have a simple closed-form formula.
- Before the widespread use of computers, people used statistical tables that show precomputed values of the cdf for different choices of (n, p) and x.

$$F(x) = P(X \leq x) = \sum_{k=0}^{x} \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}$$

		p									
n	x	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
2	0	0.9025	0.8100	0.7225	0.6400	0.5625	0.4900	0.4225	0.3600	0.3025	0.2500
	1	0.9975	0.9900	0.9775	0.9600	0.9375	0.9100	0.8775	0.8400	0.7975	0.7500
	2	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
3	0	0.8574	0.7290	0.6141	0.5120	0.4219	0.3430	0.2746	0.2160	0.1664	0.1250
	1	0.9928	0.9720	0.9392	0.8960	0.8438	0.7840	0.7182	0.6480	0.5748	0.5000
	2	0.9999	0.9990	0.9966	0.9920	0.9844	0.9730	0.9571	0.9360	0.9089	0.8750
	3	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
4	0	0.8145	0.6561	0.5220	0.4096	0.3164	0.2401	0.1785	0.1296	0.0915	0.0625
	1	0.9860	0.9477	0.8905	0.8192	0.7383	0.6517	0.5630	0.4752	0.3910	0.3125
	2	0.9995	0.9963	0.9880	0.9728	0.9492	0.9163	0.8735	0.8208	0.7585	0.6875
	3	1.0000	0.9999	0.9995	0.9984	0.9961	0.9919	0.9850	0.9744	0.9590	0.9375
	4	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

The $F(1.75) = 0.8438$ for a binomial random variable with parameters $(n = 3, p = 0.25)$.


cdf and pdf of binomial distribution
n=3
p=0.25
x=1.75
cdf = 0.8438
0.4
0.3
0.2
0.1
0.0
0
1
2
3


cdf and pdf of binomial distribution
n=20
p=0.3
x=7.3
cdf = 0.7723
0.20
0.15
0.10
0.05
0.00
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20

Exercise 6 (simple random walk)

Problem

A particle starts at the origin, and after unit time it takes a unit step left or right with probability $\frac{1}{2}$. Find the probability of being at location i after n steps where n is an even number.

Solution

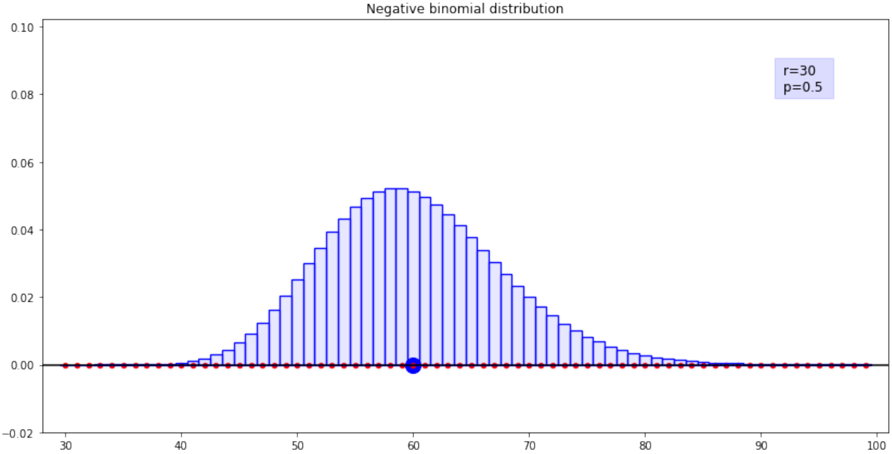
- *Think of moving right as a success.*
- *If the particle moves x steps right and $n-x$ steps left than it will be at location $i = x - (n-x) = 2x - n$.*
- *Hence, it will be at location i if and only if the number of right moves is $x = \frac{n+i}{2}$.*
- *Therefore, the probability of being at location i is equal to*

$$\begin{cases} 0 & i \text{ is odd or } |i| > n \\ f\left(\frac{n+i}{2}\right) = \binom{n}{\frac{n+i}{2}} \frac{1}{2^n} & \text{otherwise} \end{cases}.$$

Definition (Negative binomial distribution)

Let X be the number of independent Bernoulli trials with success probability p needed until exactly r successes occur. Then we say X has the **negative binomial distribution** with parameters (r, p).

- $\text{Range}(X) = \{r, r+1, \dots\}$.
- If $X(s) = x$, means there are $r-1$ successes and $x-r$ fails in the first $x-1$ trials and the $x-th$ trial is a success.
- Therefore
$$f(x) = \binom{x-1}{r-1} p^r q^{x-r}.$$
- When $r = 1$, the negative binomial distribution is the same as the geometric distribution.


Negative binomial distribution
r=30
p=0.5
0.10
0.08
0.06
0.04
0.02
0.00
−0.02
30
40
50
60
70
80
90
100

mgf of the negative binomial distribution

Using Taylor expansion, for $|z| < 1$,

$$\frac{1}{(1-z)^r} = \sum_{y=0}^{\infty} \binom{y+r-1}{r-1} z^y.$$

From the definition of mgf:

$$\begin{aligned}
M(t) = E[e^{tX}] &= \sum_{x=r}^{\infty} e^{tx} \binom{x-1}{r-1} p^r q^{x-r} \\
&= (pe^t)^r \sum_{x=r}^{\infty} \binom{x-1}{r-1} (e^t q)^{x-r} \\
&\text{(by changing } y = x - r\text{)} \\
&= (pe^t)^r \sum_{y=0}^{\infty} \binom{y+r-1}{r-1} (e^t q)^y \\
&= \left(\frac{pe^t}{1 - e^t q} \right)^r .
\end{aligned}$$

Mean and variance of the negative binomial distribution

We will spare computations, but it can be checked that

- $$E[X] = M'(0) = \frac{r}{p}.$$

- $$E[X^2] = M''(0) = \frac{r(r+q)}{p^2}.$$

- Finally,

$$\begin{aligned}\mathrm{Var}(X) &= E[X^2] - E[X]^2 \\ &= \frac{r(r+q)}{p^2} - \frac{r^2}{p^2} = \frac{rq}{p^2}.\end{aligned}$$

Poisson distribution

Definition

A discrete random variable has Poisson distribution with parameter $\lambda > 0$ if

$$\mathrm{Range}(X) = \{0, 1, 2, 3, \dots\}$$

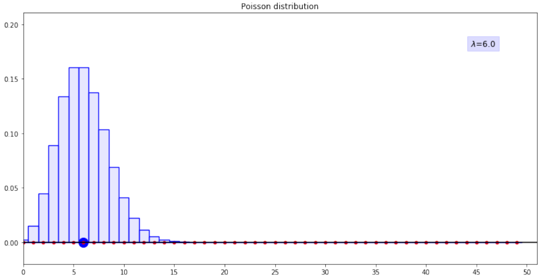
and

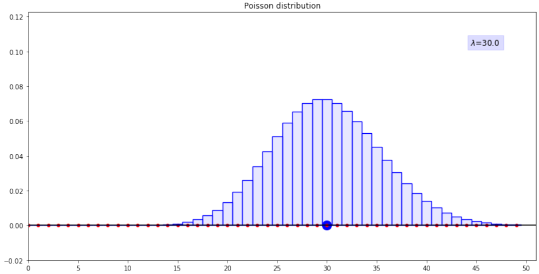
$$f(x) = \frac{e^{-\lambda}\lambda^x}{x!}, \quad x \in \mathrm{Range}(X).$$

- Notice that $f(x)$ is indeed a pmf:

$$\sum_{x \in \mathrm{Range}(X)} f(x) = \sum_{x=0}^{\infty} \frac{e^{-\lambda}\lambda^x}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = e^{-\lambda} e^{\lambda} = 1.$$

- λ is called the **expected rate**.

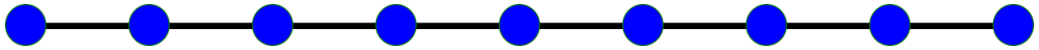

Poisson distribution
λ=6.0
0.20
0.15
0.10
0.05
0.00
0
5
10
15
20
25
30
35
40
45
50


Poisson distribution
λ=30.0
0.12
0.10
0.08
0.06
0.04
0.02
0.00
−0.02
0
5
10
15
20
25
30
35
40
45
50

Motivation for Poisson distribution

The Poisson distribution arises in the following way (made up scenario):

- Bus can arrive every $\frac{1}{n}$ hour. Sometimes it does, sometimes it does not.
- It is a very unpredictable bus but we know that there are on average λ buses arriving every hour.



- Whether a bus arrives or not at given time is independent from other times.
- Therefore, the number of buses that arrive in 1 hour is a binomial distribution with parameter $p = \frac{\lambda}{n}$.
- We know that the expect number of successes in a binomial distribution with parameters (n, p) is np and so $p = \frac{\lambda}{n}$.
- The probability of x number of buses arriving in an hour will be

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x} = \binom{n}{x} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x}, \quad x = 0, 1, 2, \ldots, n.$$

- Now imagine the bus starts arriving at more irregular times. To accommodate that we increase the n.

- When $n \to \infty$ that will correspond to the buses potentially arriving at any moment in the continuous interval.
- In that case

$$f(x) = \lim_{n\to\infty} \binom{n}{x} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x}, \quad x = 0, 1, 2, \dots.$$

- Notice that

$$\lim_{n\to\infty} \binom{n}{x} \left(\frac{\lambda}{n}\right)^x = \frac{\lambda^x}{x!} \lim_{n\to\infty} \frac{n(n-1)\cdots(n-x+1)}{n^x} = \frac{\lambda^x}{x!}.$$

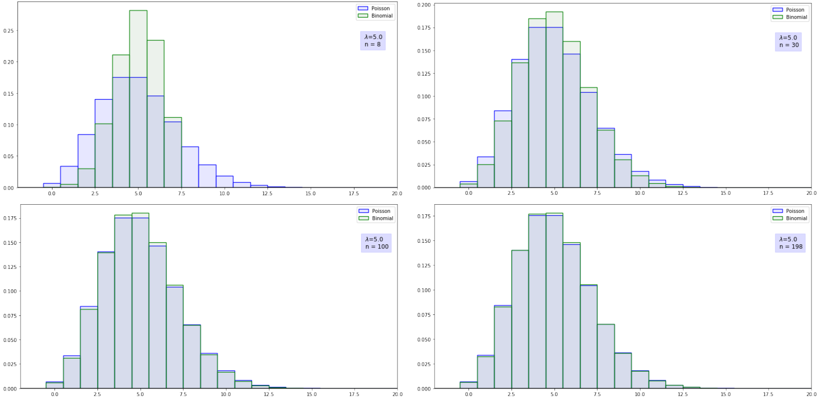
- And (you should know this from calculus)

$$\lim_{n\to\infty} \left(1 - \frac{\lambda}{n}\right)^{n-x} = e^{-\lambda}.$$

Consequently, if the buses can potentially arrive any moment within the hour, with λ buses arriving on average in every hour, then the probability that x buses will arrive in a given hour is

$$f(x) = \frac{e^{-\lambda}\lambda^x}{x!}, \quad x = 0, 1, 2, \dots.$$

Let us see that the Poisson distribution with parameter λ and the binomial distribution with parameters $(n, p = \frac{\lambda}{n})$ are close to each other by looking at their histograms when n is large.



mgf of Poisson distribution

- From the definition of mgf:

$$M(t) = E[e^{tX}] = \sum_{x=0}^{\infty} e^{tx} \frac{e^{-\lambda}\lambda^x}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{(e^t\lambda)^x}{x!} = e^{-\lambda} e^{e^t\lambda} = e^{\lambda(e^t-1)}.$$

- Therefore

$$M'(t) = \lambda e^t e^{\lambda(e^t-1)}$$
$$M''(t) = \lambda e^t e^{\lambda(e^t-1)} + \lambda^2 e^{2t} e^{\lambda(e^t-1)}$$

- Finally

$$E[X] = M'(0) = \lambda \quad \text{(as expected)}$$
$$\mathrm{Var}(X) = E[X^2] - E[X]^2 = M''(0) - \lambda^2 = \lambda + \lambda^2 - \lambda^2 = \lambda$$

Exercise 7

Problem (2.6-5 from the textbook)

Flaws in a certain type of drapery material appear on the average of one in 150 square feet. If we assume a Poisson distribution, find the probability of at most one flaw appearing in 225 square feet.

Solution

- *If the average on 150 sq ft is 1 then the average on 225 sq ft will be*

$$\lambda = \frac{225}{150} = \frac{3}{2}.$$

- *We want to find*

$$f(0) + f(1) = \frac{e^{-\lambda}\lambda^0}{0!} + \frac{e^{-\lambda}\lambda^1}{1!} = \frac{5}{2}e^{-\frac{3}{2}} \approx 0.5578.$$