Pascal Distribution

A geometric distribution models

- The number of times you roll a die until you get a one.
- The number of trips you make a trip with a car until something fails.
- The number of days until you an accident happens at work...

A Pascal distribution (also known as the negative binomial distribution) is similar, modeling

- The number of times you roll a die until you get r ones
- The number of times you make a trip with a car until r things fail (and it's time to buy a new car)
- The number of days until r accidents happen at work (and you're promoted as per the Peter principle)

Not surprisingly, the Pascal distribution is closely related to a geometric distribution.

Definitions:

- Uniform Distribution: The probability of each valid outcome is the same.
- Geometric Distribution: The number of Bernoulli trials until you get a success
- Pascal Distribution: The number of Bernoulli trials until you get r successes

Distribution	description	pdf	mgf	mean	variance
Bernoulli trial	flip a coin obtain m heads	$p^m q^{1-m}$	$q + p/z$	p	p(1-p)
Binomial	flip n coins obtain m heads	$\binom{n}{m} p^m q^{n-m}$	$(q + p/z)^n$	np	np(1-p)
Hyper Geometric	Bernoulli trial without replacement	$\frac{\binom{A}{x}\binom{B}{n-x}}{\binom{A+B}{n}}$			
Uniform range = (a,b)	toss an n-sided die	$1/n \quad a \le m \le b$ $0 \quad otherwise$	$\left(\frac{1}{n}\right)\left(\frac{1+z+z^2+...+z^{n-1}}{z^b}\right)$	$\left(\frac{a+b}{2}\right)$	$\left(\frac{(b-a+1)^2-1}{12}\right)$
Geometric	Bernoulli until 1st success	$pq^{x-1} \quad x = 1, 2, 3$ $0 \quad otherwise$	$\left(\frac{p}{z-q}\right)$	$\left(\frac{1}{p}\right)$	$\left(\frac{q}{p^2}\right)$
Pascal	Bernoulli until rth success	$\binom{x-1}{r-1} p^r q^{x-r}$	$\left(\frac{p}{z-q}\right)^r$	$\left(\frac{r}{p}\right)$	$\left(\frac{rq}{p^2}\right)$

Source: Wikipedia

Pascal Distribution

A Pascal distribution is one where you conduct a Bernoulli trial (think: flip a coin) until you get k successes. To find the pdf, assume you toss a coin with the probability of heads being p. The probability of getting r heads on nth flip is:

On the nth flip, you must get a heads:

- x=n: f() = p

On the previous n-1 flips, you got r-1 heads. This is a binomial distribution:

- $f(x) = \binom{x-1}{r-1} p^{r-1} q^{(x-1)-(r-1)}$

Since these both must happen, you get

$$\boxed{f(x|r,p) = \binom{x-1}{r-1} p^r q^{x-r}}$$

Moment Generating Function, Mean, and Variance

A Pascal distribution is the same as r geometric distibutions (the sum of r geometric distributions). In the time-domain, this results in the pdf being the convolution of the pdf's of geometric distributions. In the z-domain, it is the product.

The moment generating function for a geometric distribution is

$$\psi(z) = \left(\frac{p}{z-q}\right)$$

The moment generating function of r geometric distributions (i.e. a Pascal distribution) is thus

$$\psi(z) = \left(\frac{p}{z-q}\right)^r$$

With this, we can find the moments, mean, and variance.

m0: Zeroth Moment

To be a valid probability distribution, the total area (i.e. the zeroth moment) must be one

$$m_0 = \psi(z=1) = 1$$

$$\left.\left(\frac{p}{z-q}\right)^r\right|_{z=1} = \left(\frac{p}{p}\right)^r = 1$$

This is a valid moment generating funciton.

1st Moment: mean

The mean is the 1st moment

$$\mu = m_1 = -\psi'(z=1)$$

$$m_1 = -\frac{d}{dz}\left(\left(\frac{p}{z-q}\right)^r\right)_{z=1} = -\left(r\left(\frac{p}{z-q}\right)^{r-1}\left(\frac{-p}{(z-q)^2}\right)\right)_{z=1}$$

$$= -\left(r\left(\frac{p}{p}\right)^{r-1}\left(\frac{-p}{p^2}\right)\right)$$

$$= \left(\frac{r}{p}\right)$$

The mean of a Pascal distrbution is

$$\mu = \left(\frac{r}{p}\right)$$

2nd Moment

The second moment is

$$m_2 = \psi''(z=1)$$

$$m_2 = \frac{d^2}{dz^2}\left(\left(\frac{p}{z-q}\right)^r\right)_{z=1}$$

$$= \frac{d}{dz}\left(r\left(\frac{p}{z-q}\right)^{r-1}\left(\frac{-p}{(z-q)^2}\right)\right)_{z=1}$$

$$= \frac{d}{dz}\left(\frac{-rp^r}{(z-q)^{r+1}}\right)_{z=1}$$

$$= \left(\frac{r[r+1]p^r(z-q)^r}{(z-q)^{2(r+1)}}\right)_{z=1}$$

$$= \left(\frac{r(r+1)p^{2r}}{p^{2r+2}}\right) = \left(\frac{r(r+1)}{p^2}\right)$$

Variance

The variance is

$$\sigma^2 = m_2 - m_1^2$$

$$= \left(\frac{r(r+1)}{p^2}\right) - \left(\frac{r}{p}\right)^2 = \left(\frac{r}{p^2}\right)$$

which again is off by q. The actual variance is

$$\sigma^2 = \frac{rq}{p^2}$$

Examples of Pascal Distributions:

Example 1: Determine the pdf of

- Rolling a die until you get a 1 or 2 (p = 1/3)
- The number of times you do the dishes until someone notices (p = 1/3)
- The number of parties you go to until you catch COVID-19 (assume p = 1/3)

Solution: This is a geometric distribution

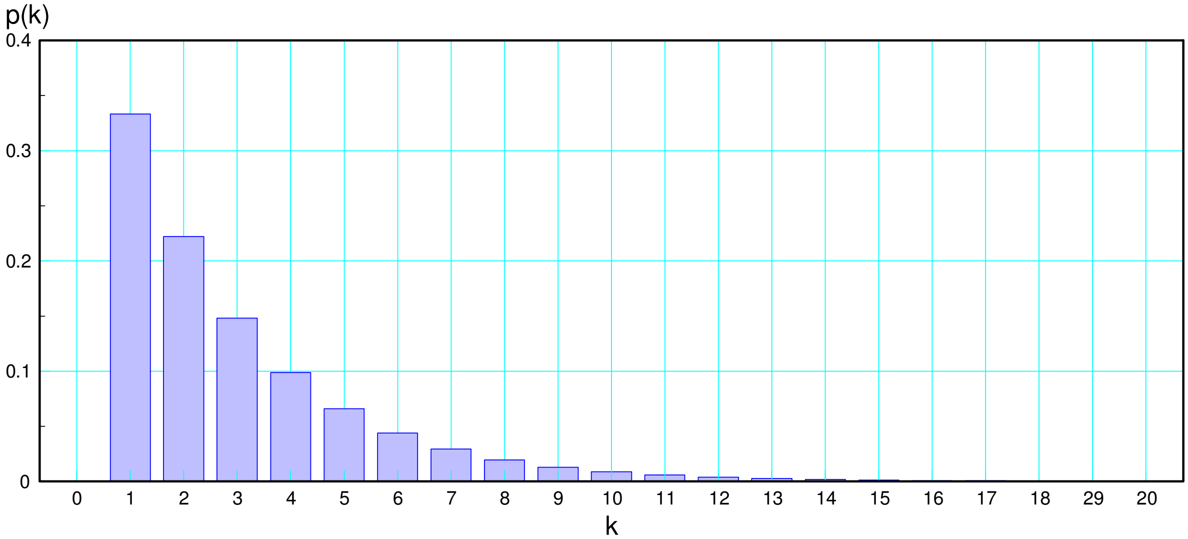
$$p(k) = \left(\frac{1}{3}\right)\left(\frac{2}{3}\right)^{k-1} u(k-1)$$

Matlab Code:

```
k = [1:30]';
A = (1/3) * (2/3).^(k-1) .* (k-1 >= 0);
k = [0;k];
A = [0;A];

bar(k,A)
```

note: 0 is added at the beginning of each array to account for k=0 (needed in convolution).



pdf for the number of rolls (k) until you roll a 1 or 2 (geometric distribution)

Example 2: Determine the pdf for

- The number of times you roll a 6-sided die until you roll a 1 or 2 twice (p = 1/3)
- The number of times you do the dishes until two people notice (p = 1/3)
- The number of parties you go to until you have two exposes to COVID-19 (assume p = 1/3)

Solution: This is a Pascal distribution.

Option 1: Convolution in matlab. We have the pdf for r = 1 from before. Use convolution to repeat the event

```
A2 = conv(A,A);
[k(1:21), A2(1:21)]

        k          A2
          0          0
     1.0000          0
     2.0000     0.1111
     3.0000     0.1481
     4.0000     0.1481
     5.0000     0.1317
     6.0000     0.1097
     7.0000     0.0878
     8.0000     0.0683
     9.0000     0.0520
    10.0000     0.0390
    11.0000     0.0289
```

Option 2: Use the Pascal distribution formula

$$p(k) = \binom{k-1}{r-1} p^r q^{k-r}$$

In Matlab:

```
p = 1/3;
q = 2/3;
r = 2;
B = zeros(21,1);
for i=3:20
   k = i-1;
   B(i) = NchooseM(k-1, r-1) * p^r * q^(k-r) .* (k >= 0);
end

k = [0:20]';
[k(1:21),A2(1:21),B(1:21)]

        k        conv    formula
          0          0          0
     1.0000          0          0
     2.0000     0.1111     0.1111
     3.0000     0.1481     0.1481
     4.0000     0.1481     0.1481
     5.0000     0.1317     0.1317
     6.0000     0.1097     0.1097
     7.0000     0.0878     0.0878
     8.0000     0.0683     0.0683
```

Note: B(k+1) is offset by one since k starts counting at zero (B(2) is actually B(k=1))

The two match up.

Option 3: z-transforms

The moment-generating function (i.e. z-transform) for a geometric distribution is

$$\psi(z) = \left(\frac{p}{z-q}\right)$$

Doing this twice gives

$$\psi(z) = \left(\frac{p}{z-q}\right)^2 = \left(\frac{1/3}{z-2/3}\right)^2$$

Take the inverse z-transform. From a table of z-transforms:

$$\left(\frac{z}{(z-a)^2}\right) \leftrightarrow \left(\frac{k}{1!}\right) a^{k-1}$$

so

$$\left(\frac{1/3}{z-2/3}\right)^2 = \left(\frac{1}{9z}\right)\left(\frac{z}{(z-2/3)^2}\right) \rightarrow \left(\frac{1}{9z}\right) k \left(\frac{2}{3}\right)^{k-1} u(k)$$

1/z means delay by one

$$p(k) = \left(\frac{1}{9}\right) (k-1) \left(\frac{2}{3}\right)^{k-2} u(k-1)$$

Checking in Matlab

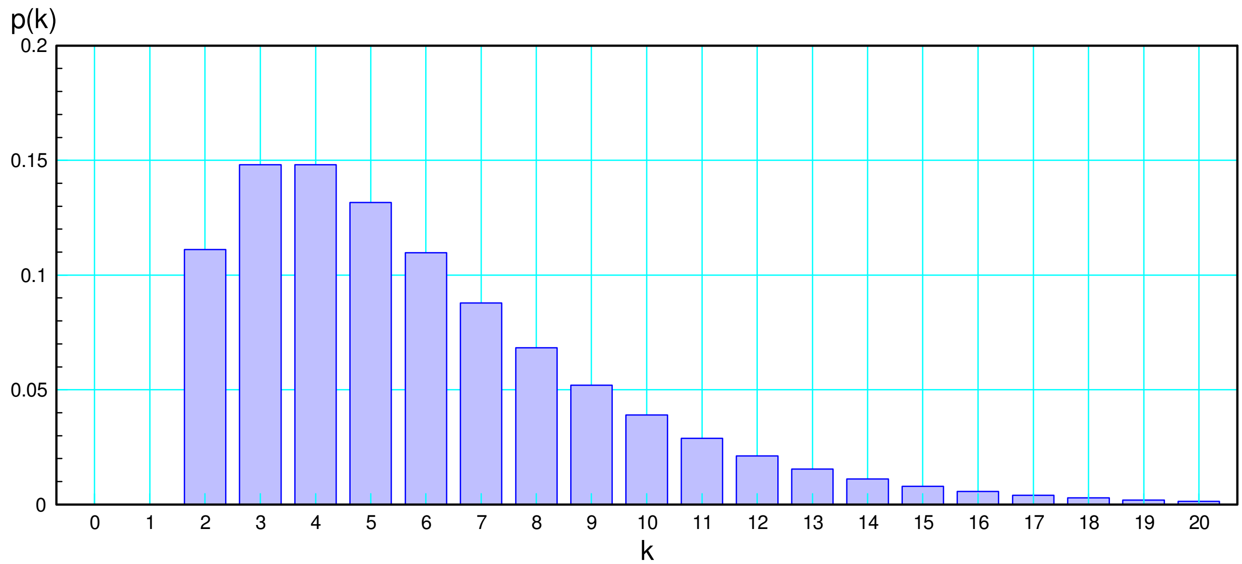
```
p = 1/3;
q = 2/3;
r = 2;

C = zeros(21,1);
for i=3:21
    k = i-1;
    C(i) = (1/9)*(k-1)* ( (2/3)^(k-2) );
    end

k = [0:20]';
[k(1:21),A2(1:21),B(1:21), C(1:21)]
```

k	conv	formula	z-trans
0	0	0	0
1.0000	0	0	0
2.0000	0.1111	0.1111	0.1111
3.0000	0.1481	0.1481	0.1481
4.0000	0.1481	0.1481	0.1481
5.0000	0.1317	0.1317	0.1317
6.0000	0.1097	0.1097	0.1097
7.0000	0.0878	0.0878	0.0878
8.0000	0.0683	0.0683	0.0683
9.0000	0.0520	0.0520	0.0520

All three methods are equivalent



Example 2:

Let

- A be the number of times you roll an 8-sided die until you roll a 1 (p = 1/8)
- B be the number of times you roll an 8-sided die until you roll a 1 or 2 (p = 2/8)
- C be the number of times you roll an 8-sided die until you roll a 1, 2, or 3 (p = 3/8)

Determine the pdf for A + B + C

Solution using Matlab and Convolutions:

A, B, and C are geometric distributions

$$A(k) = \left(\tfrac{1}{8}\right)\left(\tfrac{7}{8}\right)^{k-1} u(k-1)$$

$$B(k) = \left(\tfrac{2}{8}\right)\left(\tfrac{6}{8}\right)^{k-1} u(k-1)$$

$$C(k) = \left(\tfrac{3}{8}\right)\left(\tfrac{5}{8}\right)^{k-1} u(k-1)$$

Use Matlab to convolve the three together

```
A = (1/8) * (7/8).^(k-1) .* (k-1 >= 0);
B = (2/8) * (6/8).^(k-1) .* (k-1 >= 0);
C = (3/8) * (5/8).^(k-1) .* (k-1 >= 0);
```

```
[k,A,B,C]

         k           A           B           C
         0           0           0           0
    1.0000      0.1250      0.2500      0.3750
    2.0000      0.1094      0.1875      0.2344
    3.0000      0.0957      0.1406      0.1465
    4.0000      0.0837      0.1055      0.0916
    5.0000      0.0733      0.0791      0.0572
    6.0000      0.0641      0.0593      0.0358
    7.0000      0.0561      0.0445      0.0224
    8.0000      0.0491      0.0334      0.0140
    9.0000      0.0430      0.0250      0.0087
   10.0000      0.0376      0.0188      0.0055
```

Now convolve them

```
AB = conv(A,B);
ABC = conv(AB,C);
[k(1:21),ABC(1:21)]

     k        p(k)
     0           0
     1           0
     2           0
     3      0.0117
     4      0.0264
     5      0.0397
     6      0.0501
     7      0.0572
     8      0.0611
     9      0.0625
    10      0.0619
```

Option 2: z-transforms

The z-transforms are

$$A(z) = \left(\frac{1/8}{z-7/8}\right)$$

$$B(z) = \left(\frac{2/8}{z-6/8}\right)$$

$$C(z) = \left(\frac{3/8}{z-5/8}\right)$$

The z-transform for the sum of the three is

$$Y(z) = A(z) \cdot B(z) \cdot C(z)$$

$$Y(z) = \left(\frac{1/8}{z-7/8}\right)\left(\frac{2/8}{z-6/8}\right)\left(\frac{3/8}{z-5/8}\right)$$

This isn't in the table of z-transforms, so use partial fraction expansion

$$Y(z) = \left(\frac{1/8}{z-7/8}\right)\left(\frac{2/8}{z-6/8}\right)\left(\frac{3/8}{z-5/8}\right)$$

$$Y(z) = \left(\frac{0.375}{z-7/8}\right) + \left(\frac{-0.75}{z-6/8}\right) + \left(\frac{0.375}{z-5/8}\right)$$

This also isn't in the table of z-transforms, so multiply both sides by z

$$zY = \left(\frac{0.375z}{z-7/8}\right) + \left(\frac{-0.75z}{z-6/8}\right) + \left(\frac{0.375z}{z-5/8}\right)$$

giving

$$z \cdot y(k) = \left(0.375\left(\frac{7}{8}\right)^k - 0.75\left(\frac{6}{8}\right)^k + 0.375\left(\frac{5}{8}\right)^k\right)u(k)$$

Divide by z (delay by 1)

$$y(k) = \left(0.375\left(\frac{7}{8}\right)^{k-1} - 0.75\left(\frac{6}{8}\right)^{k-1} + 0.375\left(\frac{5}{8}\right)^{k-1}\right)u(k-1)$$

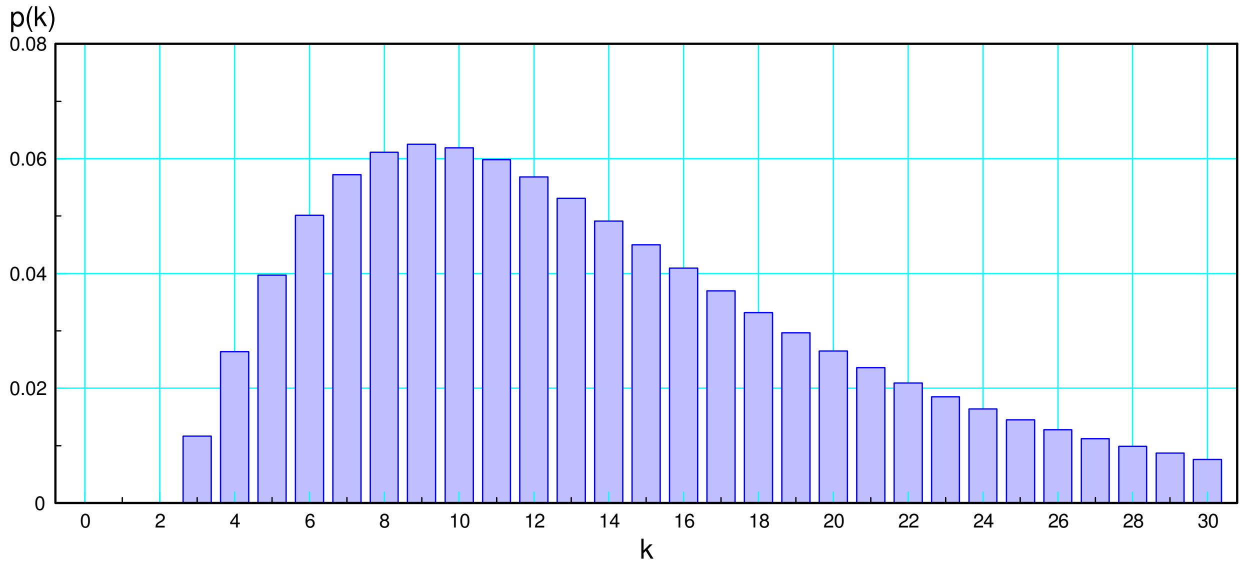
Checking against the results for convolution:

```
k = [0:31]';
Y = 0*k;
Y = ( 0.375*(7/8).^(k-1) - 0.75*(6/8).^(k-1) + 0.375*(5/8).^(k-1) ) .* (k-1>0);

[k(1:21),ABC(1:21),Y(1:21)]

   k       conv      z-trans
    0         0         0
    1         0         0
    2         0         0
    3    0.0117    0.0117
    4    0.0264    0.0264
    5    0.0397    0.0397
    6    0.0501    0.0501
    7    0.0572    0.0572
    8    0.0611    0.0611
    9    0.0625    0.0625
   10    0.0619    0.0619
   11    0.0598    0.0598
```

Either method is valid: they give you the same results.



pdf for the number of rolls of an 8-sided die until you roll a 1, then a 1 or 2, then a 1, 2, or 3

Example 4: Find the cdf

Assume the pdf of a distribution is

$$\psi(z) = \left(\frac{0.2}{z-0.8}\right)^2$$

(i.e. it's a Pascal distribution with r = 1, p = 0.2). Determine the cdf.

Solution: The cdf is the integral of the pdf. In the z-domain, this is equivalent to multiplying by $\left(\frac{z}{z-1}\right)$

$$cdf = \left(\frac{0.2}{z-0.8}\right)^2 \left(\frac{z}{z-1}\right)$$

taking the inverse z-transform

$$cdf = \left(\frac{A}{z-1}\right) + \left(\frac{B}{(z-0.8)^2}\right) + \left(\frac{C}{z-0.8}\right)$$

Using the cover-up method, A and B are

$$A = \left(\left(\frac{0.2}{z-0.8}\right)^2 \left(\frac{z}{-}\right)\right)_{z=1} = 1$$

$$B = \left(\left(\frac{0.2^2}{-}\right)\left(\frac{z}{z-1}\right)\right)_{z=0.8} = -0.16$$

C is found from

$$C = \frac{d}{dz}\left(\left(\frac{0.2}{-}\right)^2\left(\frac{z}{z-1}\right)\right)_{z=0.8}$$

$$C = 0.04\left(\frac{(z-1)-z}{(z-1)^2}\right)_{z=0.8}$$

$$C = -1$$

giving

$$cdf = \left(\frac{1}{z-1}\right) + \left(\frac{-0.16}{(z-0.8)^2}\right) + \left(\frac{-1}{z-0.8}\right)$$

This isn't in the table of z-transforms, so multiply by z

$$z\ cdf = \left(\frac{z}{z-1}\right) + \left(\frac{-016z}{(z-0.8)^2}\right) + \left(\frac{-z}{z-0.8}\right)$$

Take the inverse z transform

$$z\ cdf = \left(1 - 0.16k\ (0.8)^k - (0.8)^k\right)u(k)$$

divide by z (delay by 1)

$$cdf = \left(1 - 0.16(k-1)\ (0.8)^{k-1} - (0.8)^{k-1}\right)u(k-1)$$