# Chapter 3: Common Families of Distribution

October 3, 2021

## Exercise 3.1

There are $N_1 - N_0 + 1$ numbers, therefore $P(x = n) = \frac{1}{N_1 - N_0 + 1}$.

$EX = \frac{N_1 + N_0}{2}$ which is just the midpoint.

Let $b = N_1, a = N_0$

$$\begin{aligned}
\text{Var}X &= EX^2 - (EX)^2 \\
&= \frac{1}{b-a+1}\sum_a^b x^2 - (EX)^2 \\
&= \frac{1}{b-a+1}\left[\sum_1^b x^2 - \sum_1^{a-1} x^2\right] - (EX)^2 \\
&= \frac{1}{b-a+1}\left[\frac{b(b+1)(2b+1)}{6} - \frac{(a-1)a(2a-1)}{6}\right] - \frac{(b+a)^2}{4} \\
&= \frac{2b(b+1)(2b+1) - 2(a-1)a(2a-1) - 3(b-a+1)(b+a)^2}{12(b-a+1)} \\
&= \frac{2b(b-a+1+a)(2b+1) + 2a(b-a+1-b)(2a-1) - 3(b-a+1)(b+a)^2}{12(b-a+1)} \\
&= \frac{2b(b-a+1)(2b+1) + 2a(b-a+1)(2a-1) - 3(b-a+1)(b+a)^2 - 4ab(b-a+1)}{12(b-a+1)} \\
&= \frac{2b(2b+1) + 2a(2a-1) - 3(b+a)^2 - 4ab}{12} \\
&= \frac{a^2 + b^2 - 2ab + 2b - 2a}{12} \\
&= \frac{(b-a)(b-a+2)}{12} \\
&= \frac{(N_1 - N_0)(N_1 - N_0 + 2)}{12}
\end{aligned}$$

## Exercise 3.2

(a) Let $X$ be the number of defective part in $K$ samples and $M$ be the total defective parts in 100 parts. Then

$$P(X=0|M>5) = \frac{\binom{100-M}{K}}{\binom{100}{K}}$$

is the probability of accepting a defective product given $M > 5$. To bound $K$, we can set $M = 6$ since defect parts becomes more prevalent which increases the chance for them to be sampled, setting $M = 6$ maximizes the false positive rate $P(X = 0|M)$.

Then

$$P(X=0|M=6) = \frac{\binom{94}{K}}{\binom{100}{K}} < 0.1$$

Solving for $K$ numerically (polynomial of the 5th power), we get $K > 31$. We can choose $K = 32$.

(b) The false positive rate is now

$$P(X \leq 1|M=6) = P(X=0|M=6) + P(X=1|M=6) = \frac{\binom{94}{K}}{\binom{100}{K}} + \frac{\binom{6}{1}\binom{94}{K-1}}{\binom{100}{K}} < 0.1$$

Solving for $K$ numerically (same as above except there's an additional term $1 + \frac{6K}{95-K}$), We get $K > 50.24$ which means $K = 51$.

## Exercise 3.4

(a) Without eliminating the wrong key, every trial is independent with probability of $\frac{1}{n}$ for succeeding. Let $X$ be the number of tries before succeeding.

$$P(X=k) = \left(1 - \frac{1}{n}\right)^{k-1} \frac{1}{n}$$

It is geometric distribution therefore $EX = \frac{1}{1/n} = n$

(b) By eliminating the wrong key, at $k$-th trial, we will be selecting from $n - k + 1$ remaining keys, the success probability is $\frac{1}{n-k+1}$.

$$P(X=k) = \frac{1}{n-k+1} \prod_{i=1}^{k-1} \left(1 - \frac{1}{n-i+1}\right)$$

Then

$$
\begin{aligned}
EX &= \sum_{x=1}^{n} xP(x) \\
&= \sum_{k=1}^{n} \frac{k}{n-k+1} \prod_{i=1}^{k-1} \left(1 - \frac{1}{n-i+1}\right) \\
&= \sum_{k=1}^{n} \frac{k}{n-k+1} \prod_{i=1}^{k-1} \frac{n-i}{n-i+1} \\
&= \sum_{k=1}^{n} \frac{k}{n-k+1} \frac{n-k+1}{n} \\
&= \sum_{k=1}^{n} \frac{k}{n} \\
&= \frac{n+1}{2}
\end{aligned}
$$

# Exercise 3.5

We want to first assume that if new drug has the same effectiveness of the old drug and see how often would we observe the result of 85 effective cases out of 100 to make sure the experiment result is not a coincident. (Or how often standard drug can produce the same trial result as the new drug).

Let $X$ be the number of effective cases out of 100. Then $X \sim \text{Binomial}(100, 0.8)$.

$$P(X \geq 85) = \sum_{k=85}^{100} \binom{100}{85} 0.8^{85} 0.2^{15} \approx 0.1285$$

So there are about $13\%$ chance that the standard drug can produce the same result. We can't conclude the new drug is better.

# Exercise 3.6

(a) We can use binomial distribution $\text{Binomial}(2000, 0.01)$. Then

$$P(X = k) = \binom{2000}{k} 0.01^{k} 0.99^{2000-k}$$

(b)

$$P(X < 100) = \sum_{k=0}^{99} \binom{2000}{k} 0.01^{k} 0.99^{2000-k}$$

(c) A conservative rule to check if a normal approximation is good is $\min(np, n(1-p)) = 20 \geq 5$. Therefore we can approximate it with $\mathcal{N}(\mu = 20, \sigma = \sqrt{20(0.99)})$.

$$P(X < 100) = P\left(\frac{X - 20}{\sqrt{20(0.99)}} \leq (\frac{100 - 20}{\sqrt{20(0.99)}}\right) = P(Z < 17.98) \approx 1$$

# Exercise 3.7

Let $X$ be the number of chocolate chips in the cookie and $X \sim \text{Poisson}(\lambda)$. Then

$$P(X \geq 2) = 1 - P(X = 0) - P(X = 1) = 1 - e^{-\lambda} - \lambda e^{-\lambda} > 0.99$$

We get $\lambda \approx 6.64$.

# Exercise 3.8

(a) Let $X$ be the number of people who choose theatre 1. Then when $X \leq N$ will be the event theatre 1 not turning away customers and $1000 - X \leq N$ will be when theatre 2 not turning away customers. We will find the reverse: $N$ such that the probability of both theatre not turning away customers is greater than 0.99 . $P(1000 - N \leq X \leq N) > 0.99)$

The binomial distribution for $X$ is

$$P(X = k) = \binom{1000}{k} 0.5^{1000}$$

From $1000 - N \leq X \leq N$, we get $N \geq 500$ (If total seats of two theatres is less than number of customers, one of them will certainly turn away customers).

We have

$$0.5^{1000} \sum_{1000-N}^{N} \binom{1000}{k} > 0.99, \quad N \geq 500$$

```
from scipy.special import comb
def p(N):
  return sum([comb(1000, i, exact=True) for i in range(1000-N, N+1)])
      / (2**1000)
```

We get $p(540) \approx 0.9896$ and $p(541) \approx 0.9913$. Therefore we can take $N = 541$.

(b) Since $X \sim \text{Binomial}(1000, 1/2)$. We can approximate with a normal distributed $Y \sim \mathcal{N}(1000 \times 0.5, \sqrt(1000 \times 0.5 \times 0.5)) = \mathcal{N}(500, \sqrt{250})$

$$P(1000 - N \leq X \leq N) \approx P(1000 - N \leq Y \leq N) = P(-\frac{N - 500}{\sqrt{250}} \leq Z \leq \frac{N - 500}{\sqrt{250}}) > 0.99$$

This means we need to look for a $z$ score where the probability is greater than 0.995 by symmetry. So $z \geq 2.58$. So $N \geq 2.58(\sqrt{250}) + 500 = 540.8 \approx 541$

# Exercise 3.20

(a)

$$\mathbf{E}X = \int_0^\infty x f(x) dx = \frac{2}{\sqrt{2\pi}} \int_0^\infty x e^{-x^2/2} dx = -\frac{2}{\sqrt{2\pi}} e^{-x^2/2}|_0^\infty = \frac{2}{\sqrt{2\pi}}$$

$$\mathbf{E}X^2 = \int_0^\infty x^2 f(x) dx = \frac{2}{\sqrt{2\pi}} \int_0^\infty x^2 e^{-x^2/2} dx$$

To calculate the integral, use integration by part $dg = xe^{-x^2/2}dx$, $f = x$. Therefore $g = -e^{-x^2/2}$ and $df = dx$.

$$\int_0^\infty x^2 e^{-x^2/2} dx = -xe^{-x^2/2}|_0^\infty + \int_0^\infty e^{-x^2/2} dx = \sqrt{\frac{\pi}{2}}$$

So

$$EX^2 = \frac{2}{\sqrt{2\pi}} \sqrt{\frac{\pi}{2}} = 1$$

Therefore $\mathrm{Var}X = \mathbf{E}X^2 - (\mathbf{E}X)^2 = 1 - \left(\frac{2}{\sqrt{2\pi}}\right)^2 = 1 - \frac{2}{\pi}$

(b) Let $y = sx^2$, by change of variable,

$$f_Y(y) = f_X(x(y)) \left|\frac{dx}{dy}\right| = \frac{1}{\sqrt{2\pi s}} y^{1/2} e^{-\frac{y}{2s}}$$

Now we compare it again the gamma distribution

$$f_Y(y|\alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta^\alpha} y^{\alpha-1} e^{-y/\beta}$$

First we conclude $\beta = 2s$, $\alpha = 1/2$. Then we have $\Gamma(\alpha)\beta^\alpha = \Gamma(1/2)\sqrt{2s} = \sqrt{2\pi s}$ which is consistent with the above. Therefore the change of variable is $y = \frac{\beta}{2}x^2$, and $Y \sim gamma(\alpha = 1/2, \beta > 0)$