5. It is thought that human influenza viruses originate in birds. It is quite possible that, several years ago, a human influenza pandemic was averted by slaughtering 1.5 million chickens brought to market in Hong Kong. Because it is impossible to test each chicken individually, such decisions are based on samples. Suppose that a boy has already died of a bird flu virus apparently contracted from a chicken. Several diseased chickens have already been identified. The health officials would prefer to err on the side of caution and destroy all chickens that might be infected; the farmers do not want this to happen unless it is absolutely necessary. Suppose that both the farmers and the health officals agree that all chickens should be destroyed if more than 2 percent of the population is diseased. A random sample of $n = 1000$ chickens reveals 40 diseased chickens.

   (a) Let $X_i = 1$ if chicken i is diseased and $X_i = 0$ if it is not. Assume that $X_1, \ldots, X_n \sim P$. To what family of probability distributions does P belong? What population parameter indexes this family? Use this parameter to state formulas for $\mu = EX_i$ and $\sigma^2 = \text{Var}\, X_i$.

   (b) State appropriate null and alternative hypotheses from the perspective of the health officials.

$\mu = EX_i = p$

b) $H_0: \mu \geq 0.02$ (p)

$H_1: \mu < 0.02$

$\text{Var}\, X_i = p \cdot (1-p) = \sigma^2$

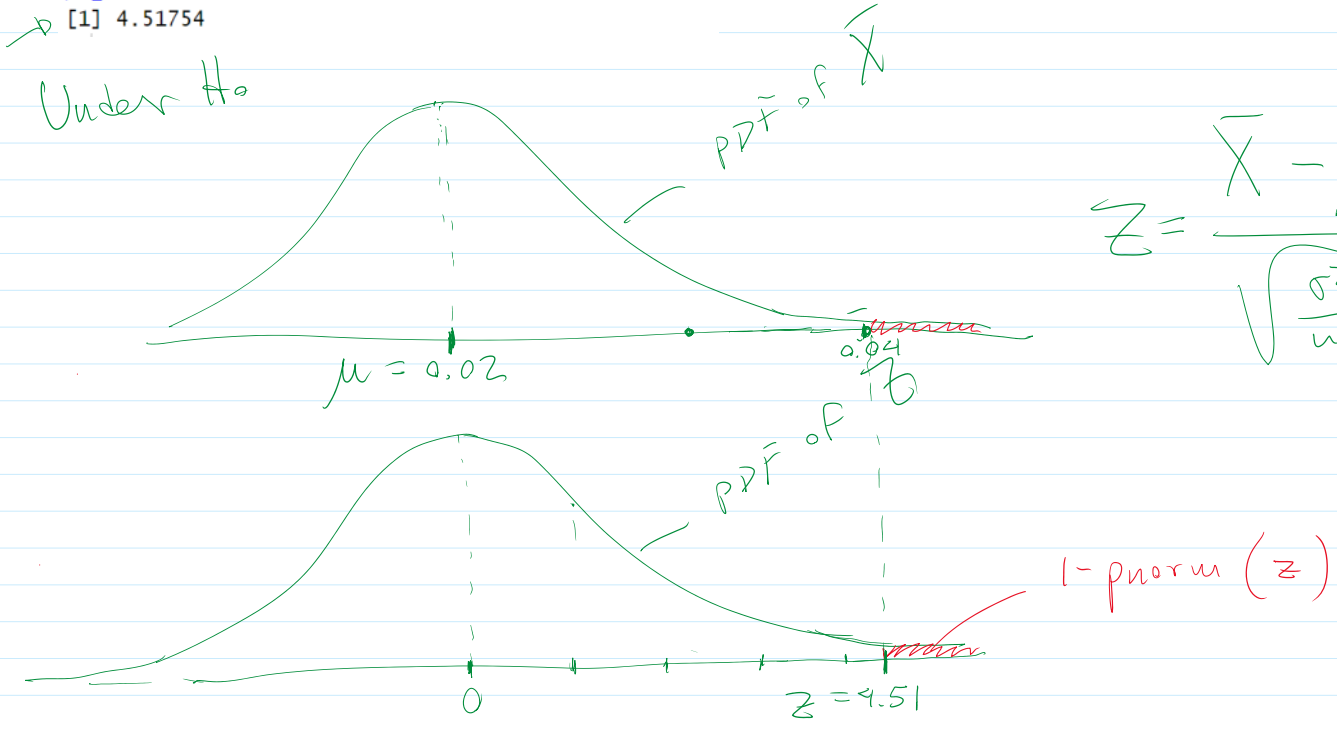
Under H_0:

$\sigma^2 = 0.02 \cdot (1 - 0.02)$

Perform the test

$$z = \frac{\bar{x} - \mu_0}{\sqrt{\frac{\sigma^2}{n}}} = \frac{\frac{40}{1000} - 0.02}{\sqrt{\frac{0.02 \cdot (1-0.02)}{1000}}}$$

```
> z = (40/1000 - 0.02)/(sqrt(0.02*(1-0.02)/1000))
> z
[1] 4.51754
```



$$z = \frac{\bar{X} - \mu_0}{\sqrt{\frac{\sigma^2}{n}}}$$

c) Farmers' perspective

$H_0: \mu \leq 0.02$ ✱

$H_0: \mu \leq 0.02$

$H_1: \mu > 0.02$

```
> 1-pnorm(z)
[1] 3.128118e-06
```

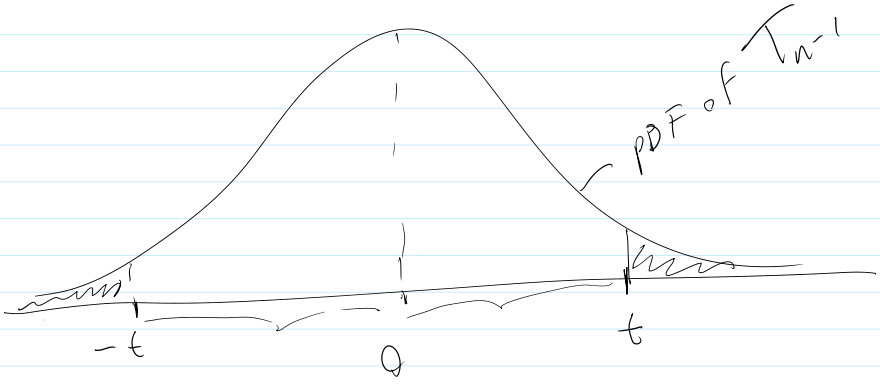
0.0000031

We reject H_0

We conclude that more than 2% of chickens (in Hong Kong) are diseased.

What if: $H_0: \mu = 10$

$H_1: \mu \neq 10$



p-value = 2*(1-pt(abs(t), n-1))

Simulation-based approach to inference

- Bootstrap samples.
  - Random samples with replacement obtained from the original sample
    - All of them have the same size.

These are not the only simulated sample we use.

We also can use
- permutations
- draws. (Bernoulli)

(See 03-23-23 lab)

```
> 2*(1-pt(t, n-1))
[1] 0.03755959
```