MAT 2377
Probability and Statistics for Engineers

Practice Set

P. Boily (uOttawa)

Winter 2021

Based on course notes by Rafał Kulik

Q36. In a clinical study, volunteers are tested for a gene that has been found to increase the risk for a particular disease. The probability that the person carries a gene is 0.1.

a) What is the probability that 4 or more people will have to be tested in order to detect 1 person with the gene?

b) How many people are expected to be tested in order to detect 1 person with the gene?

c) How many people are expected to be tested in order to detect 2 people with the gene?

Solution: if X is the number of tests in order with the last one being the 1^{st} success (gene detection), then X has geometric distribution with $p = 0.1$

a) In this case, we want to evaluate

$$P(X \geq 4) = \sum_{k=4}^{\infty} (1-p)^{k-1} p = p \frac{(1-p)^3}{1-(1-p)} = (1-p)^3 = 0.729.$$

b) By definition, $\mathrm{E}[X] = 1/p = 1/0.1 = 10.$

c) We can think of this procedure as splitting the patients into 2 groups, randomly, and testing each of the groups one by one. It takes on average 10 tests in order to detect 1 gene in either one of the groups, so 20 for the two combined groups.

Q38. Use `R` to generate a sample from a binomial distribution and from a Poisson distribution (select parameters as you wish).

Use `R` to compute the sample means and sample variances. Compare these values to population means and population variances.

Solution: samples from the desired distributions are generated by the functions `rbinom(n,size,prob)` and `rpois(n,lambda)`.

We simulate `n=1000` values for each sample. **WARNING:** in the course, n is used to represent the number of trials; in `R`, n is used to represent the sample size, and the number of trials is represented by the parameter `size`. It's not what I would have chosen.

For the binomial distribution X, we use `size=20` and `prob=0.2`; for the Poisson distribution Y, we use `lambda=8.5`.

The true underlying means and variances are

$$\mathrm{E}[X] = 20(0.2) = 4, \quad \mathrm{Var}[X] = 20(0.2)(0.8) = 3.2$$
$$\mathrm{E}[Y] = \mathrm{Var}[Y] = 8.5.$$

The samples and estimated values agree with the theoretical ones:

```
X=rbinom(1000,20,0.2);
mean(X); var(X);
[1] 4.022 [1] 3.0745901
```

```
Y=rpois(1000,8.5);
mean(Y); var(Y);
[1] 8.388 [1] 9.090547
```

Agreement is a broad term to use. What does it mean, in this context? We'll revisit this at a later stage.