Probabilities & Data
Week 6 – Special Discrete Distributions & Bayesian Inference

DR NICK FEWSTER-YOUNG

Topics to be covered

- Sampling without replacement
- Hypergeometric Distribution
- Binomial Approximation to Hypergeometric Distribution
- Extended Hypergeometric
- Modelling
- Discrete Predicative Distribution by Bayesian Methods

Sampling Without Replacement

- The models which we have encountered so far, Bernoulli, Binomial, Possion and Trinomial are by no means the only statistical models which are encounter.
- Suppose that we sample without replacement from a finite population instead of the previous ideal of replacing then this leads to a different story.
- The process becomes non-stationary and the probabilities depend on the various possibilities of the previous trials.

Example

- Consider a deck of 52 cards which is well shuffled.
- Example: The conditional probability of drawing a heart on the fifth draw is dependent on the previous 4 draws.
- Why? The probability that you draw four straight hearts could happen! Then the fifth as a chance of 9/48. Maybe 2 hearts are drawn, or maybe none!
- In this case, this is not a Bernoulli process because we are not replacing the cards. This relates to the process having memory!

- This leads to the hypergeometric model!

Hypergeometric Distribution

This is the statistical model used to represent situation of sampling without replacement.

Let

- r number of “successes” in the sample,
- n the number of trials (sample size),
- R the number of “successes” in the entire population and
- N represent the number of items in the entire population.

Therefore the distribution is given by

$$P(r = r_1 \mid n, R, N) = \frac{\binom{R}{r}\binom{N-R}{n-r}}{\binom{N}{n}}$$

Hypergeometric Distribution

$$P(r = r_1 \mid n, R, N) = \frac{\binom{R}{r}\binom{N-R}{n-r}}{\binom{N}{n}}$$

- The denominator represents the number of different possible samples of size n that can be drawn without replacement from a population of size N.
- The numerator represents the number of these possible samples that include exactly r ”successes” and n-r ”failures”.
- The first term in the numerator represents the choice/combination of ”successes” r from R, while the second term considers the case of the failures.

Example (Back to cards - Poker)

Consider our 52 cards and let r be the number of hearts in a five-card hand. By using the hypergeometric distribution find the probability you get dealt, 0-5 hearts in your hand.

$$P(r = 0\,|n = 5, R = 13, N = 52) = \frac{\binom{13}{0}\binom{39}{5}}{\binom{52}{5}} = 0.2215$$

$$P(r = 1\,|n = 5, R = 13, N = 52) = \frac{\binom{13}{1}\binom{39}{5}}{\binom{52}{5}} = 0.4114$$

$$P(r = 2\,|n = 5, R = 13, N = 52) = \frac{\binom{13}{2}\binom{39}{5}}{\binom{52}{5}} = 0.2743$$

$$P(r = 3\,|n = 5, R = 13, N = 52) = \frac{\binom{13}{3}\binom{39}{5}}{\binom{52}{5}} = 0.0815$$

$$P(r = 4\,|n = 5, R = 13, N = 52) = \frac{\binom{13}{4}\binom{39}{5}}{\binom{52}{5}} = 0.0107$$

$$P(r = 5\,|n = 5, R = 13, N = 52) = \frac{\binom{13}{5}\binom{39}{5}}{\binom{52}{5}} = 0.0005$$

Exercise (Neighborhoods)

Question: Suppose a neighborhood consists of 40 families with yearly incomes of at least \$52,000 and 60 families with yearly incomes below \$52,000. What is the probability of observing exactly five families with incomes of at least \$52,000 in a random sample of 10 families.

Solution: By using the hypergeometric distribution with r = 5, n = 10, R = 40, N = 60:

$$P(r = 5\ |n = 10, R = 40, N = 100) = \frac{\binom{40}{5}\binom{60}{5}}{\binom{100}{10}} = 0.21.$$

How close is this to Binomial?

- Suppose we add replacement to test how close these probabilities are!
- We are picking 5 families out of 10 families where the probability of their income being greater than \$52,000 is 40/100 = 0.4.
- Thus, we have

$$P(r = 5 \,|n = 10, p = 0.4) = \binom{10}{5}(0.4)^5(0.6)^5 = 0.2007.$$

➢ This is close to the without replacement answer…how???

- When the sample size is small compare to the population size then the Binomial approximates the Hypergeometric distribution with

$$p = \frac{R}{N}.$$

Remember the Poisson Distribution?

- Ideally, we have the same result and transitive result because we can use the Poisson Distribution to approximate the Binomial distribution.
- Isn't that nice ☺
- The assumptions are very similar: we need a small

$$p = \frac{R}{N}$$

and a large N. Perfecto ☺

Exercise: Use the Poisson distribution to check the previous answer also for the neighborhoods.

Extending the Hypergeometric Distribution

- We can extend the distribution to analyzing more different types of “successes” like when we bootstrapped to the trinomial distribution!

- Suppose there are K different types of “successes” or “fails” and denote each number of each “success” or “fail” by k_i. Assume that the last one k_K is the fails.
- Thus this gives us

$$P(r_1 = k_1, \dots, r_K = k_K, | n, R_1, \dots, R_K) = \frac{\binom{R_1}{k_1}\binom{R_2}{k_2}\dots\binom{R_K}{k_K}}{\binom{N}{n}}$$

where $\sum_{i=1}^{K} k_i = n$ and $\sum_{i=1}^{K} R_i = N$.

Poker Cards

What is the probability of obtaining 1 jack of any suit, 2 kings of any suit, 1 ace of any suit and 1 queen of any suit in a five card hand.

Solution: We use the extended hypergeometric function:

$$P(r_1 = 1, r_2 = 2, r_3 = 1, r_4 = 1, |5,13,13,13,13) = \frac{\binom{13}{1}\binom{13}{2}\binom{13}{1}\binom{13}{1}}{\binom{52}{5}} =$$

Conditional Probability Revisited

- Conditional probability is one of the most important concepts is statistics and especially Bayes' theorem.
- To get us back in the mindset, let's consider the following example:

Auto-car Dealership

Example: The owner of a car dealership has just hired a new salesman. Based on the past experience with the salesmen and on a personal interview with the new employee, the owner judges that the odds are 4 to 1 against him being a **great (sells 1 car every 2 days)** salesman but the odds are even that he could a **good (1 every 4 days)** salesmen. This implies that the probability that he is a great salesmen is

$$\frac{1}{5} = 0.2$$

And the probability that he is a good salesmen 0.5. It follows that the only other option is he useless or **poor (1 every 8 days)** with probability 0.3.

The owner has a prior distribution for car sales based on the performance of a salesman given

$$P\left(\lambda = \frac{1}{2}\right) = 0.2, P\left(\lambda = \frac{1}{4}\right) = 0.5, P\left(\lambda = \frac{1}{8}\right) = 0.3.$$

Car dealership

1. What distribution or process is this modelling for the conditional probabilities of selling 10 cars per 24 days given a type of salesmen?

2. Calculate the probabilities of each type of salesman sells 10 cars in 24 days?

3. Compute the posterior probabilities using Bayes theorem, which are

$$P(\lambda_i | y = 10).$$

Predicative Distributions

Bayes' theorem is used to revise probabilities concerning a random variable θon the basis of new sample information that is represented by y. Inferences or conclusions concerning θ may then be based on the posterior distribution of θ determined by Bayes Theorem.

In some scenarios, you may be interested in making **predications** about the sample result before it is observed.

Your uncertainty is expressed in terms of the sample outcome, y in the form of a probability distribution, which is called the predicative distribution. The possible outcomes in the sample are denoted by $y_1, \dots, y_k$ together with their probabilities.

Motivation

- Usually, a predicative probability is important is decision making for the future and betting for example which depends on the sample outcome.
- For instance, consider a simple bet of flipping a coin. There are two outcomes, Heads or Tails.
- Another, might be in a production line, where k is the number of defective items on the line. Which ones will be defective in the future of a shipping containing 20 items.
- Another, the owner of a auto dealership may be interested not in the general rate at which a particular salesman sells cars, rather in the exact number of cars sold by that salesman in a particular month. So, perhaps the owner must order some cars from the factory in advance and in order to decide how many cars to order needs to determine his predicative probabilities for the number for cars that will be sold.

An very important application!

Helping identify whether or not to sample and more generally, deciding the size of the sample to take.

Soooo…..how do we calculate and determine predictive probabilities??? The basis comes from Bayes' theorem and we could work them out by hand.

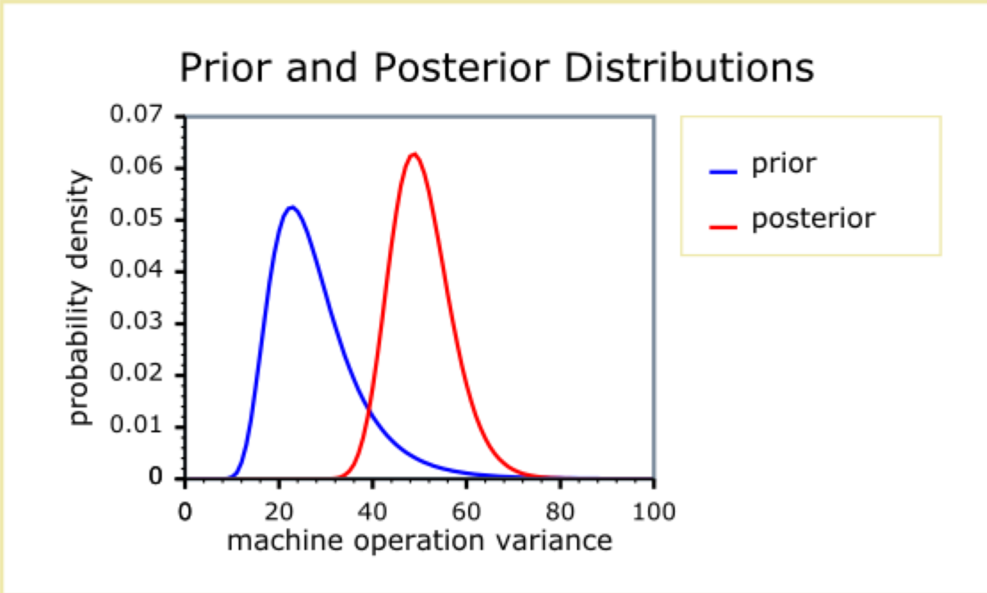
Bayes' Theorem:

$$P\left(\theta_j \middle| \, y_k\right) = \frac{P(y_k|\theta_j)P\left(\theta_j\right)}{\sum_{i=1}^{J} P(y_k|\theta_i)P(\theta_i)},$$

Where θ_j represent a value of the random variable θ and y_k represents new sample information.

The prior distribution

We already know our θ_j and the probabilities associated, they form our prior distribution because it represents the state of uncertainty about θ prior to a particular sample.



Posterior Distribution

- The values of θ together with the conditional probabilities $P(\theta_j|y_k)$ constitute the conditional probability distribution of θ_j given knowledge about the sample results y_k.

- It represents the state of uncertainty about θ after observing the sample result y_k.

- In applying Bayes' theorem, the ***likelihoods*** are determined from the conditional distributions of y_k given the different values of θ_j.

- The prior and posterior distributions may be used to make inferences or decisions concerning θ before and after observing the sample result y_k.

Predicative Distribution

We want to form a predicative distribution, this is our $P(y = y_k)$ where we don't know them!

So how do we do it??

Bayes' theorem, it all comes down to a rearrangement: see that

$$P(y_k \cap \theta_i) = P(y_k|\theta_i)\, P(\theta_i)$$

and recall that the $\sum_{i=1}^{J} P(y_k \cap \theta_i) = P(y_k)$. Thus, we have

$$P(y_k) = \sum_{i=1}^{J} P(y_k|\, \theta_i) P(\theta_i).$$

Predicative Distribution

Suppose we calculate the predicative the distribution, the best way to store its vital information is in a table form.

Let’s attempt an example involving the auto-dealership where the sample result is 10 sales in 24 days, and the predicative probability for this sample result is 0.042.

This means at the time the salesman is hired, the probability that he will sell exactly 10 cars in his first 24 days is 0.042.

Car-dealership

Recall the prior probabilities and the computed posterior distribution from the exercise earlier:

Let λ be ”rate” a salesmen sells a car. The owner has a prior distribution for car sales based on the performance of a salesman given

$$P\left(\lambda = \frac{1}{2}\right) = 0.2, P\left(\lambda = \frac{1}{4}\right) = 0.5, P\left(\lambda = \frac{1}{8}\right) = 0.3.$$

The posterior distribution of being “a great salesmen”, ”a good salesmen” , “a poor salesmen” is

$$P\left(\lambda = \frac{1}{2} \mid y = 10\right) = 0.501$$
$$P\left(\lambda = \frac{1}{4} \mid y = 10\right) = 0.493$$
$$P\left(\lambda = \frac{1}{8} \mid y = 10\right) = 0.006.$$

Question

What is the probability that in the next 24 days, he will again sell exactly 10 cars?

By using the posterior probabilities and noting the likelihoods determined from the Poisson distribution:

$$P(y = 10 \mid \lambda = \frac{1}{2}) = 0.1048$$

$$P(y = 10 \mid \lambda = \frac{1}{4}) = 0.0413$$

$$P(y = 10 \mid \lambda = \frac{1}{8}) = 0.0008$$

We can calculate the predicative probability of $y = 10$ by using Bayes' Rule.

The predicative probability

The predicative probability of $y = 10$ is

$$P(y = 10) = P(y = 10 \mid \lambda = \tfrac{1}{2})P(\lambda = \tfrac{1}{2}) + P(y = 10 \mid \lambda = \tfrac{1}{4})P\left(\lambda = \tfrac{1}{4}\right)$$

$$+ \qquad P(y = 10 \mid \lambda = \tfrac{1}{8})P(\lambda = \tfrac{1}{8}).$$

$$= 0.1048{\times}0.501 + 0.0431{\times}0.493 + 0.0008{\times}0.006$$

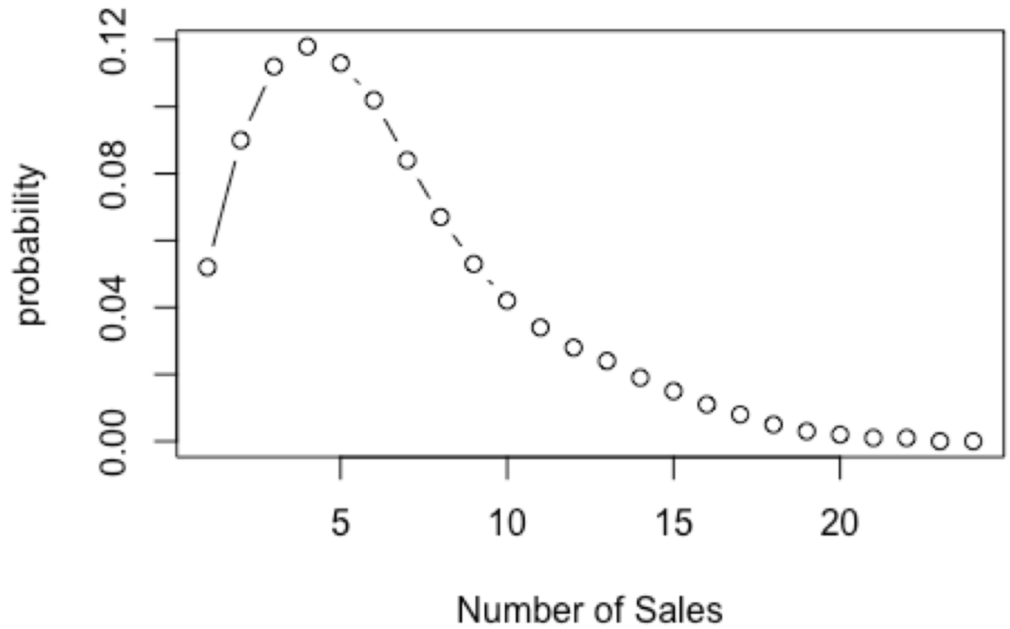
$$= 0.073.$$

Note that this is just the calculation for the predicative event that $y = 10$, that is the salesmen would sell 10 cars in 24 days. So, we could continue our calculations and find the probabilities / distribution for $y = 0, \ldots, 24$.

The table for the distribution y

y	$P(y = y_k)$	y	$P(y = y_k)$
0	0.016	13	0.024
1	0.052	14	0.019
2	0.090	15	0.015
3	0.112	16	0.011
4	0.118	17	0.008
5	0.113	18	0.005
6	0.102	19	0.003
7	0.084	20	0.002
8	0.067	21	0.001
9	0.053	22	0.001
10	0.042	23	0.000
11	0.034	24	0.000
12	0.028		

Probability Density Plot



Sensitivity – small perbutations

After the new salesmen sells 10 cars in his first 24 days, the predicative probability that he sell exactly 10 more cars in the next 24 days is 0.073. But the owner believes that his ability or rating as a salesmen should be increased. Therefore, he changes the prior distribution by placing a higher probability on being “great” and in turn decreasing the probability of being “poor” salesmen when making 10 sales in 24 days.

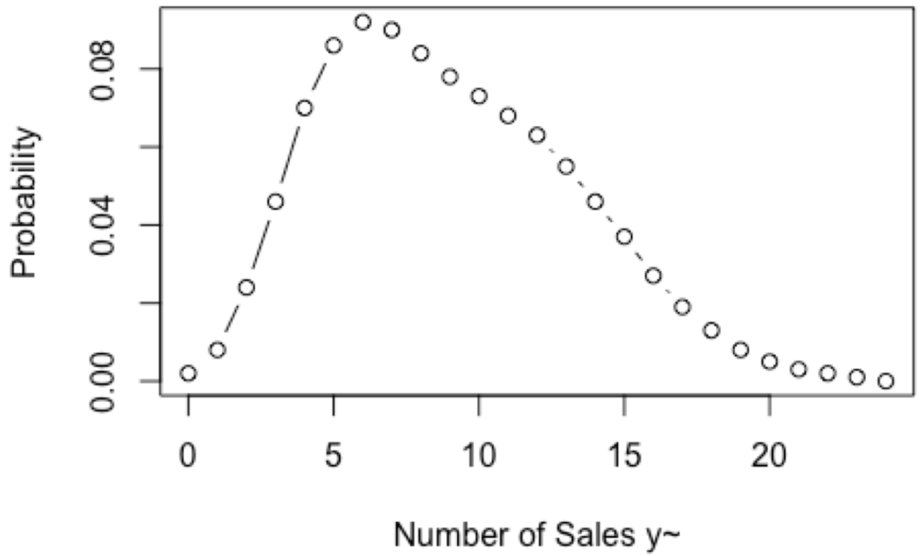
How does this change the predicative distribution?

Alter the prior probabilities by 0.05 and 0.10, and see what happens!

New Predicative Model

$\tilde{y}$	$P(\tilde{y} = \tilde{y}_k)$	$\tilde{y}$	$P(\tilde{y} = \tilde{y}_k)$
0	0.002	13	0.055
1	0.008	14	0.046
2	0.0024	15	0.037
3	0.046	16	0.027
4	0.070	17	0.019
5	0.086	18	0.013
6	0.092	19	0.008
7	0.090	20	0.005
8	0.084	21	0.003
9	0.078	22	0.002
10	0.073	23	0.001
11	0.068	24	0.000
12	0.063		

New Probability Plot



Conclusions

- As compared to the previous predicative distribution, we see that low values of y are less likely to occur and high values of y are more likely to occur.
- This demonstrates the dependence of the predicative distribution on the prior distribution.
- Note the likelihoods are identical in both cases, just the prior knowledge is different.
- This is the being of Bayesian Analysis and is very important in decision-making scenarios.
- The next time, we will encounter Bayesian is its theory for continuous distributions.

Have a good break!

- Have a good break and see you in 2 weeks!