Statistics for Data Science

Lecture 2

Counting, Bayes Theorem, Mass and Density Functions

Ken Cameron

Bayes Theorem, Mass and Density Functions

- Counting Approaches
- Bayes Theorem
- Mass/Density Functions

c. If $A \subset B$, then $P(A) <= P(B)$

- Did you manage to prove it?

- Answer:

  If $A \subset B$ then $A \cap B = A$.

  From (a) $P(B \cap A^c) = P(B) – P(A \cap B)$ we therefore have,

$$0 <= P(B \cap A^c) = P(B) – P(A)$$

The Frequentist Approach

- Statisticians can turn counting into a very sophisticated and powerful process.
- The challenge is that the counting can be subject to many restrictions.
- The answer is to break a problem into simpler problems that are easier to count and then combine those using our known rules.

Separate Tasks

- Theorem: If a job consists of k separate tasks, the i^{th} of which can be done in n_i ways, i = 1, …, k, then the entire job can be done in

$$n_1 \times n_2 \times n_3 \times \ldots \times n_k \text{ ways.}$$

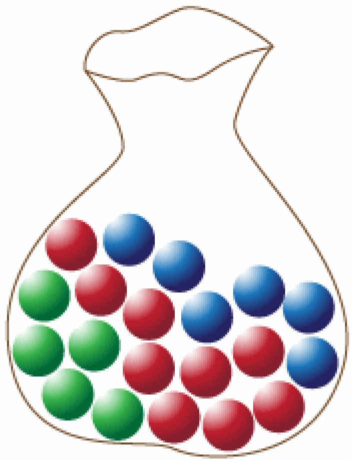
- Proof: Sufficient to prove for k = 2. The proof is just a matter of careful counting. The first task can be done in n1 ways, and for each of these ways we have n2 choices for the second choice. So we can perform the job in

$$(1 \times n_2) + (1 \times n_2) + (1 \times n_2) + \ldots + (1 \times n_2) = n_1 \times n_2$$

^^^^ n_1 ways ^^^^

ways.

Replacement



- We need to consider if counting is done with our without replacement.
- Consider picking your lottery numbers.
  - Would it make a difference if you could choose the same number twice?
  - Your first choice is from 59. Your second is from 58. So there are 59x58 choices available. (3422)
  - With replacement it's 59x59 choices. (3481)

Ordered or not?

- Does the order of the choices matter?
  - If we treat the choice 1, 2 as equal to 2, 1, then there as half as any combinations (assuming without replacement).
- Two useful notations:
  - n! n-factorial. It's the product of all the positive integers between 1 and n. i.e. n x (n-1) x (n-2) x … x 3 x 2 x 1.
  - And the symbol $\binom{n}{r} = \frac{n!}{r!(n-r!)}$

Ordered or not?

- Six ordered without replacement 59!/53!
- Six ordered with replacement 59^6
- Six unordered without replacement 59!/(6!53!)
- Six unordered with replacement …

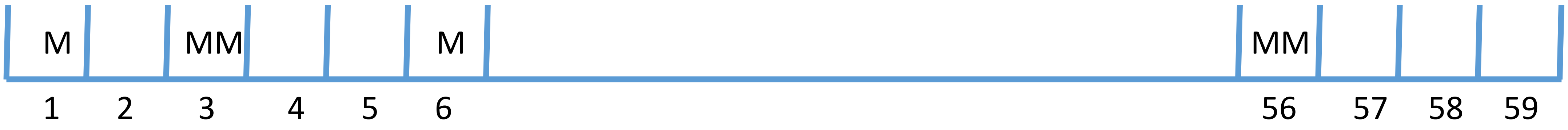
Unordered with Replacement

- The hardest case to count.

- Let’s try visualising it:

  - Think of each of our 59 defining a bin and we’re going to distribute 6 markers between the bins.

  - The answer is then the number of ways we can put the 6 markers in the 59 bins.

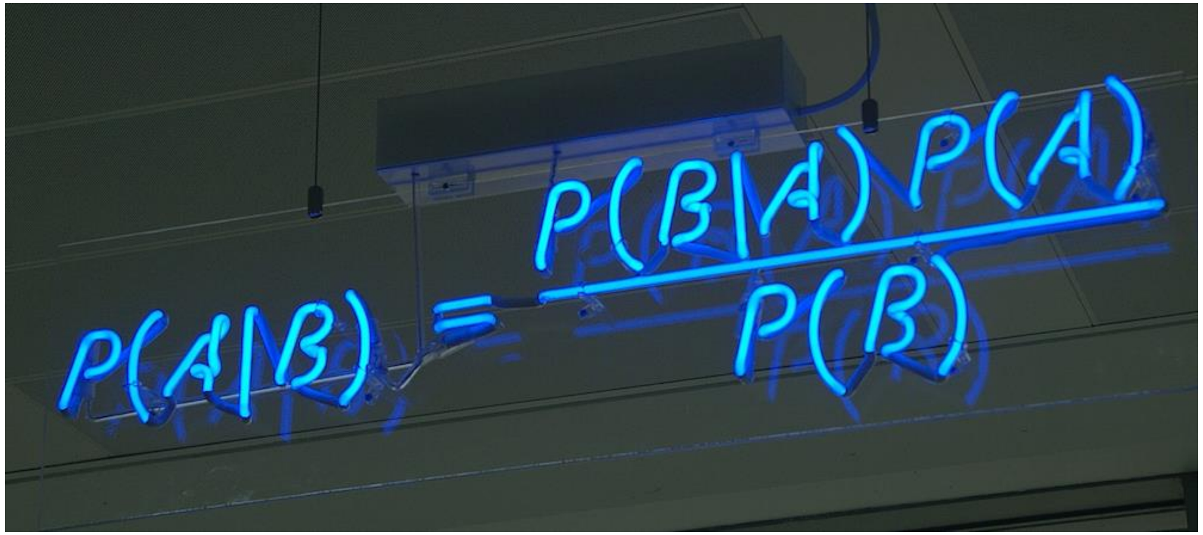


Unordered with Replacement

- We only need to consider the arrangement of the bin walls and markers.
- We can ignore the outer walls.
- So only need to consider the 58 walls and 6 markers.
- We have 58 + 6 = 64 objects that can be arranged in 64! ways.
- However we have to eliminate the redundant orderings so divide by 6! and 58!

$$\frac{64!}{6!\,58!}$$

Beyond Counting



- Bayes Theorem/Law.

- This rule simple is one of the most powerful tools when manipulating probabilities.

Bayes' Law

$$p(y|x) = \frac{p(x|y)p(y)}{p(x)}$$

- This is the Bayesian principle that allows us to incorporate our *prior beliefs* into our mathematical statements
- In particular we call
  - $p(y)$ the prior probability
  - $p(y|x)$ the posterior probability
  - $p(x|y)$ the likelihood
  - $p(x)$ the evidence density

Example

- A realistic hypothetical

  - The police are stopping vehicles at random and testing them for drink driving with a breathalyser
  - 1 in 1000 of the people on the road are drunk
  - The device has been rigorously tested and it has shown that it has 95% accuracy
  - They stop someone, and the test is positive

- What is the probability that they are drunk?

Example

- Denote drunk as D, positive test as $+$

- We want to know $p(D|+)$, let's use Bayes:

$$p(D|+) = \frac{p(+|D)p(D)}{p(+)}$$

- From the question
  - $p(+|D) = 0.95$
  - $p(D) = 0.001$

- Can calculate $p(+)$

$$p(+) = p(+, D) + p(+, D')$$
$$p(+) = p(+|D)p(D) + p(+|D')p(D')$$

Complement, i.e. sober

Example

- Continuing…

$$p(+) = p(+|D)p(D) + p(+|D')p(D')$$

- From question:
  - $p(+|D') = 0.05$
  - $p(D') = 1 - p(D) = 0.999$
- So, $p(+) = 0.95 \times 0.001 + 0.05 \times 0.999 = 0.0509$
- Hence

$$p(D|+) = \frac{0.95 \times 0.001}{0.0509} \approx 0.019$$

Example

- Less than 2%!
- The fact is that 95% accuracy is not very good given how rare (0.1%) it is for someone to be drunk in the first place, the false positives hugely outweigh the true positives
- Bayes' Law lets us put that accuracy into context

- Even if the accuracy was better the outcome would still be low.
  - So how do you get to 'beyond a reasonable doubt'?

Distribution Functions

- For every random variable X, we have an associated cumulative distribution function of X. (cdf)

$$F_X(x) = P_X(X <= x), \text{ for all } x.$$

- All cdf satisfy certain properties:
  - $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$
  - F(x) is a non-decreasing function
  - F(x) is right continuous. i.e. for every x_0, $\lim_{x \downarrow x_0} F(x) = F(x_0)$

Distribution Functions Example

- Example: Tossing three coins. X = number of heads observed.

$$Fx(x) = \begin{cases} 0 & \text{if } -\infty <= x < 0 \\ 1/8 & \text{if } 0 <= x < 1 \\ 1/2 & \text{if } 1 <= x < 2 \\ 7/8 & \text{if } 2 <= x < 3 \\ 1 & \text{if } 3 <= x < -\infty \end{cases}$$

Density and Mass Functions

- Associated with a random variable X and its cdf F(x)is another function:

- Either, the probability density function (pdf)
  - For continuous random variables.

- Or, the probability mass function (pmf)
  - For discrete random variables.

Probability Mass Functions

- $f_x(x) = P(X = x)$ for all X
- Note that $P(X = x)$ is the size of the jump in the cdf at x.
- We can use the pmf to calculate probabilities. For example,

$$f_x(x) = P(X = x) = \begin{cases} (1-p)^{k-1}p \; for \; x = 1,2, \ldots \\ 0 \qquad\qquad otherwise \end{cases}$$

$$P(a <= X <= b) = \sum_{k=a}^{b} fx(k) = \sum_{k=a}^{b} (1-p)^{k-1}p.$$

Probability Density Functions

- A probability density function of a continuous random variable X satisfies:

$$F_x(x) = \int_{-\infty}^{\infty} f_x(t)dt \ \ for\ all\ x.$$

$$\frac{d}{dx}F_x(x) = fx(x)$$