

Probability with Discrete Variables

Ben Goodrich

January 31, 2022

Review of Last Week

- For Frequentists like Fisher and Neyman, probability is needed to describe the consequences of explicit randomization:
 - Random sampling of units from a much larger population
 - Random assignment of units to treatment or control groups
 - Random measurement or other physical error
- Supervised learning usually randomizes which observations are put in the training or testing set but mostly does not use probability to describe the implications of doing so and thus cannot characterize uncertainty
- Bayesians disagree sharply with both of these schools of thought
 - Probability is fundamental for science because we need to **describe all forms of uncertainty**, whether due to overt randomization or not

really important element of science - but frequentists can't do!

Random Variables (R.V.)

- A function is a rule that UNIQUELY maps each element of an input set to some element of an output set, e.g. e^x maps real numbers (\mathbb{R}) to non-negative real numbers (\mathbb{R}_+) we're never talking about complex numbers - A L W A Y S real numbers! (in this class)
- A random variable is a FUNCTION from the sample space, Ω , to some subset of \mathbb{R} with a probability-based rule
- If Ω is discrete with a finite number of elements, then we can simply enumerate an equivalent number of probabilities to define a random variable

```
die <- 1:6 # Omega in this case
```

```
roll <- sample(die, size = 1, prob = rep(1 / 6, times = 6)) # code calls a function
```

```
# roll is a realization of the function set of function where we take one element from the sample space (omega)
```

- Do not conflate a REALIZATION of a R.V. with the FUNCTION that generated it
- By convention, a capital letter, X , indicates a R.V. and its lower-case counterpart, x , indicates a realization of X

Basics of Bowling

Each “frame” in bowling starts with $n = 10$ pins & you get up to 2 rolls per frame

✗ How to Pick Up the 7 - 10 Split in Bowling 🎳



Approaching Bowling Probabilistically

- What is Ω for your first roll of a frame of bowling? 0 to 10 (0:10)
- If $b^p = x$, then $\log_b(x) = p$. Let the probability of knocking down X out of n pins be given by a form of Benford's Law: $\Pr(x \mid n) = \log_{n+2} \left(1 + \frac{1}{n+1-x} \right)$

probability of knocking down x out of n pins

```
Pr <- function(x, n = 10) ifelse(x > n, 0, log(1 + 1 / (n + 1 - x), base = n + 2))
```

```
Omega <- 0:10 # 0, 1, ..., 10
```

```
names(Omega) <- as.character(Omega)
```

```
source("bowling.R") # does the above
```

```
x <- sample(Omega, size = 1, prob = Pr(Omega)) # realization of bowling random variable
```

```
round(c(Pr(Omega), total = sum(Pr(Omega))), digits = 5)
```

```
##      0      1      2      3      4      5      6      7      8      9     10
## 0.03502 0.03836 0.04240 0.04740 0.05374 0.06203 0.07337 0.08980 0.11577 0.16317 0.27894
## total
## 1.00000
```

if your probabilities = 1, then very likely you did it right; if != 1, then very likely you did it wrong!!!

Second Roll in a Frame of Bowling

- How would you compute the probability of knocking down all the remaining pins on your second roll?
- Let X_1 and X_2 respectively be the number of pins knocked down on the first and second rolls of a frame of bowling. What function yields the probability of knocking down x_2 pins on your second roll?
- $\Pr(x_2 \mid n = 10 - x_1) = \log_{10-x_1+2} \left(1 + \frac{1}{10-x_1+1-x_2} \right) \times \mathbb{I}\{x_2 \leq 10 - x_1\}$
- $\mathbb{I}\{\cdot\}$ is an “indicator function” that equals 1 if it is true and 0 if it is false
- $\Pr(x_2 \mid n = 10 - x_1)$ is a **CONDITIONAL probability** because it **depends** (via n) **on the realization of x_1**

From Aristotelian Logic to Bivariate Probability

- In R, TRUE maps to 1 and FALSE maps to 0 when doing arithmetic operations

```
c(TRUE & TRUE, TRUE & FALSE, FALSE & FALSE)
```

```
## [1] TRUE FALSE FALSE
```

```
c(TRUE * TRUE, TRUE * FALSE, FALSE * FALSE) | & = multiplication
```

```
## [1] 1 0 0
```

- Can generalize to probabilities on the $[0, 1]$ interval to compute the probability that two (or more) propositions are true simultaneously
- \cap reads as “and”. **General Multiplication Rule:**

$$\Pr(A \cap B) = \Pr(B) \times \Pr(A | B) = \Pr(A) \times \Pr(B | A)$$

Independence

- Loosely, A and B are independent propositions if A being true or false tells us nothing about the probability that B is true (and vice versa)
- Formally, A and B are independent iff $\Pr(A|B) = \Pr(A)$ (and $\Pr(B|A) = \Pr(B)$). Thus, $\Pr(A \cap B) = \Pr(A) \times \Pr(B)$.
- Why is it reasonable to think
 - Two rolls in the same frame are not independent?
 - Two rolls in different frames are independent?
 - Rolls by two different people are independent regardless of whether they are in the same frame?
- What is the probability of obtaining a turkey (3 consecutive strikes)? $\text{prob}(\text{strike})^3$
- What is the probability of knocking down x pins on the first roll and $10 - x$ pins on the second roll? [use general multiplication rule!!](#)

Joint Probability of Two Rolls in Bowling

- How to obtain the joint probability, $\Pr(x_1 \cap x_2 \mid n = 10)$, in general?

$$\Pr(x_1 \cap x_2 \mid n = 10) = \Pr(x_1 \mid n = 10) \times \Pr(x_2 \mid n = 10 - x_1) = \log_{12} \left(1 + \frac{1}{11 - x_1} \right) \times \log_{12 - x_1} \left(1 + \frac{1}{11 - x_1 - x_2} \right) \times \mathbb{I}\{x_1 + x_2 \leq 10\}$$

```
joint_Pr <- matrix(0, nrow = 11, ncol = 11, dimnames = list(Omega, Omega))
for (x_1 in Omega) { # joint_Pr is already filled in by source("bowling.R")
  Pr_x_1 <- Pr(x_1, n = 10)
  for (x_2 in 0:(10 - x_1))
    joint_Pr[x_1 + 1, x_2 + 1] <- Pr_x_1 * Pr(x_2, n = 10 - x_1)
}
sum(joint_Pr) # that sums to 1, can do View(joint_Pr) to see it better than on the next slide
good check for loop - if sum(joint_Pr) = 1

## [1] 1
```

joint_Pr: Row is roll 1, Column is roll 2

	0	1	2	3	4	5	6	7	8	9	10
0	0.001226	0.001343	0.001485	0.00166	0.001882	0.002172	0.002569	0.003144	0.004054	0.005714	0.009767
1	0.001525	0.001685	0.001884	0.002136	0.002466	0.002916	0.003569	0.004602	0.006486	0.011087	0
2	0.00194	0.002169	0.002459	0.002839	0.003357	0.004109	0.005297	0.007466	0.012764	0	0
3	0.002541	0.002881	0.003325	0.003933	0.004814	0.006206	0.008747	0.014953	0	0	0
4	0.003451	0.003984	0.004712	0.005766	0.007434	0.010478	0.017912	0	0	0	0
5	0.004914	0.005812	0.007114	0.009171	0.012926	0.022097	0	0	0	0	0
6	0.007466	0.009138	0.01178	0.016604	0.028384	0	0	0	0	0	0
7	0.01245	0.016051	0.022623	0.038675	0	0	0	0	0	0	0
8	0.024025	0.033861	0.057886	0	0	0	0	0	0	0	0
9	0.060222	0.10295	0	0	0	0	0	0	0	0	0
10	0.278943	0	0	0	0	0	0	0	0	0	0

joint probability of two things != conditional probabilities of one thing given another thing

Aristotelian Logic to Probability of Alternatives

c(TRUE | FALSE, FALSE | FALSE, TRUE | TRUE)

[1] TRUE FALSE TRUE

c(TRUE + FALSE, FALSE + FALSE, TRUE + TRUE - TRUE * TRUE) OR = addition usually
general addition rule!!

[1] 1 0 1

- What is the probability you fail to get a strike on this frame or the next one?
- Can generalize Aristotelian logic to probabilities on the $[0, 1]$ interval to compute the probability that one of two (or more) propositions is true
- \cup is read as “or”. **General Addition Rule:**
$$\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$$
- Iff $\Pr(A \cap B) = 0$, A and B are mutually exclusive (disjoint)

What is the probability that $X_2 = 9$?

	0	1	2	3	4	5	6	7	8	9	10
0	0.001226	0.001343	0.001485	0.00166	0.001882	0.002172	0.002569	0.003144	0.004054	0.005714	0.009767
1	0.001525	0.001685	0.001884	0.002136	0.002466	0.002916	0.003569	0.004602	0.006486	0.011087	0
2	0.00194	0.002169	0.002459	0.002839	0.003357	0.004109	0.005297	0.007466	0.012764	0	0
3	0.002541	0.002881	0.003325	0.003933	0.004814	0.006206	0.008747	0.014953	0	0	0
4	0.003451	0.003984	0.004712	0.005766	0.007434	0.010478	0.017912	0	0	0	0
5	0.004914	0.005812	0.007114	0.009171	0.012926	0.022097	0	0	0	0	0
6	0.007466	0.009138	0.01178	0.016604	0.028384	0	0	0	0	0	0
7	0.01245	0.016051	0.022623	0.038675	0	0	0	0	0	0	0
8	0.024025	0.033861	0.057886	0	0	0	0	0	0	0	0
9	0.060222	0.10295	0	0	0	0	0	0	0	0	0
10	0.278943	0	0	0	0	0	0	0	0	0	0

IF you knocked down 9 pins on the second roll, you would have HAD to knock down 0 or 1 pins (which are mutually exclusive) in the first roll

uses general ADDITION rule, in the case of mutual respect - and then add together joint probabilities

Marginal Probabilities from Joint Probabilities

	0	1	2	3	4	5	6	7	8	9	10		row-sum
0	0.001226	0.001343	0.001485	0.00166	0.001882	0.002172	0.002569	0.003144	0.004054	0.005714	0.009767	-1	0.035016
1	0.001525	0.001685	0.001884	0.002136	0.002466	0.002916	0.003569	0.004602	0.006486	0.011087	0	-1	0.038356
2	0.00194	0.002169	0.002459	0.002839	0.003357	0.004109	0.005297	0.007466	0.012764	0	0	-1	0.0424
3	0.002541	0.002881	0.003325	0.003933	0.004814	0.006206	0.008747	0.014953	0	0	0	-1	0.047399
4	0.003451	0.003984	0.004712	0.005766	0.007434	0.010478	0.017912	0	0	0	0	-1	0.053737
5	0.004914	0.005812	0.007114	0.009171	0.012926	0.022097	0	0	0	0	0	-1	0.062035
6	0.007466	0.009138	0.01178	0.016604	0.028384	0	0	0	0	0	0	-1	0.073372
7	0.01245	0.016051	0.022623	0.038675	0	0	0	0	0	0	0	-1	0.0898
8	0.024025	0.033861	0.057886	0	0	0	0	0	0	0	0	-1	0.115772
9	0.060222	0.10295	0	0	0	0	0	0	0	0	0	-1	0.163171
10	0.278943	0	0	0	0	0	0	0	0	0	0	-1	0.278943
	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	
col-sum	0.398702	0.179873	0.113268	0.080783	0.061263	0.047979	0.038095	0.030165	0.023303	0.016801	0.009767	-11	1

← marginal probabilities

↑

Marginal Distribution of Second Roll in Bowling

- How to obtain $\Pr(x_2)$ irrespective of x_1 ?
- Since events in the first roll are mutually exclusive, use the simplified form of the General Addition Rule to “marginalize”:

$$\begin{aligned}\Pr(x_2) &= \sum_{x=0}^{10} \Pr(X_1 = x \cap X_2 = x_2 \mid n = 10) \\ &= \sum_{x=0}^{10} \Pr(x \mid n = 10) \times \Pr(x_2 \mid n = 10 - x)\end{aligned}$$

- This corresponds to a column-wise sum of joint_Pr
- Better notation might be $\Pr(\cancel{x_1} \cap x_2 \mid n = 10)$ better way to show it!

Marginal, Conditional, and Joint Probabilities

- To compose a joint (in this case, bivariate) probability, **MULTIPLY a marginal probability by a conditional probability**
- To decompose a joint (in this case, bivariate) probability, **ADD the relevant joint probabilities to obtain a marginal probability**
- To obtain a conditional probability, **DIVIDE the joint probability** (in this case, bivariate) **by the marginal probability of the event** that you want to condition on because

$$\Pr(A \cap B) = \Pr(B) \times \Pr(A | B) = \Pr(A) \times \Pr(B | A) \implies$$

$$\Pr(A | B) = \frac{\Pr(A) \times \Pr(B | A)}{\Pr(B)} \text{ if } \Pr(B) > 0$$

- This is Bayes' Rule
- What is an expression for $\Pr(X_1 = 3 | X_2 = 4)$ in bowling?

Conditioning on $X_2 = 4$ in Bowling

[illegible]

$\Pr(X_1 = 3 \mid X_2 = 4)$ by Simulation vs. Bayes' Rule

```
library(dplyr)
S <- 10^7 # practically infinite
tibble(x_1 = sample(0mega, size = S, replace = TRUE, prob = Pr(0mega)),
       x_2 = vapply(x_1, FUN.VALUE = integer(1), FUN = function(x) {
         # draw from conditional probability distribution given each value of x_1
         sample(0mega, size = 1, prob = Pr(0mega, n = 10 - x))
       }))) %>%
filter(x_2 == 4) %>%
summarize(simulated = mean(x_1 == 3)) %>%
mutate(exact = joint_Pr["3", "4"] / sum(joint_Pr[, "4"])) %>%
print(digits = 4)
```

```
## # A tibble: 1 × 2
##   simulated exact
##   <dbl>   <dbl>
## 1    0.0790 0.0786
```

Bayesian vs Frequentist Probability

- Bayesians generalize this by taking A to be “beliefs about whatever you do not know” and B to be whatever you do know in

$$\Pr(A|B) = \frac{\Pr(A) \times \Pr(B|A)}{\Pr(B)} \text{ if } \Pr(B) > 0$$

- Frequentists accept the validity Bayes' Rule but object to using the language of probability to describe beliefs about unknown propositions and insist that probability is a property of a process that can be defined as a limit

$$\Pr(A) = \lim_{S \uparrow \infty} \frac{\text{times that } A \text{ occurs in } S \text{ independent randomizations}}{S}$$

limits to observables

Probability that a Huge Odd Integer is Prime

- John Cook [asks](#) an interesting question: What is the probability x is prime, where x is a huge, odd integer like $1 + 10^{100,000,000}$?
- To Frequentists, x is not a random variable. It is either prime or composite and it makes no sense to say that it is “probably (not) prime”
- To Bayesians, x is either prime or composite but no one knows for sure which
- The prime number theorem implies provides a way to choose the prior probability that x is prime based on its number of digits, (d)

$$\Pr(x \text{ is prime} \mid d) = \frac{1}{d \ln 10} \approx \frac{1}{2.3 \times 10^{10}}$$

although you could double that by taking into account that x is odd

- What is the probability that $\beta > 0$ in a regression model?

Expectation of a Discrete Random Variable

```
round(Pr(Omega), digits = 5) # What is the mode, median, and expectation of X_1?
```

```
##           0           1           2           3           4           5           6           7           8           9          10
## 0.03502 0.03836 0.04240 0.04740 0.05374 0.06203 0.07337 0.08980 0.11577 0.16317 0.27894
```

- The MODE is the element of Ω with the highest probability (10 here)
- The MEDIAN is the smallest element of Ω such that AT LEAST half of the cumulative probability is less than or equal to that element (8 here)
- EXPECTATION of a discrete random variable X is defined as

$$\mathbb{E}X = \sum_{x \in \Omega} [x \times \Pr(x)] \equiv \mu$$

- Expectation is just a probability-weighted sum (7.043 here)

Practice Problems

- How would we calculate the expectation of the second roll given that $x_1 = 7$ pins were knocked down on the first roll?
- Answer: $\sum(\Omega * \Pr(\Omega, n = 10 - 7))$, which is 1.9746359
- How would we calculate the expectation of the second roll in a frame of bowling?

Marginal Expectation of Second Roll in Bowling

$$\begin{aligned}\mathbb{E}X_2 &= \sum_{x_j=0}^{10} x_j \Pr(X_2 = x_j) \\&= \sum_{x_j=0}^{10} x_j \sum_{x_i=0}^{10} \Pr(x_i \cap x_j \mid n = 10) \\&= \sum_{x_j=0}^{10} x_j \sum_{x_i=0}^{10} \Pr(x_i \mid n = 10) \Pr(x_j \mid n = 10 - x_i)\end{aligned}$$

```
Pr_X2 <- colSums(joint_Pr) # marginal probabilities of the second roll
EX2 <- sum(Omega * Pr_X2) # definition of marginal expectation
EX2
```

```
## [1] 2.008738
```

Expectations of Functions of Discrete RVs

- Let $g(X)$ be a function of a discrete random variable whose expectation is

$$\mathbb{E}g(X) = \sum_{x \in \Omega} [g(x) \times \Pr(x)] \neq g(\mathbb{E}X)$$

- It is often sensible to make a decision that maximizes EXPECTED utility:
 1. Enumerate D possible decisions $\{d_1, d_2, \dots, d_D\}$ that are under consideration
 2. Define a utility function $g(d, \dots)$ that also depends on unknown (and maybe some known) quantities
 3. Obtain / update your conditional probability distribution for all the unknowns given all the knowns using Bayes' Rule
 4. Evaluate $\mathbb{E}g(d, \dots)$ for each of the D decisions
 5. Choose the decision that has the highest value in (4)

Blackjack

- How can you play a card game, such as blackjack, if you can't see all the cards?
- Make the decision to **Hit** or **Stand** (or possibly **Double down** or **SPlit**) to maximize expected profit. That collection of decisions forms a strategy. If we assume there are an infinite number of decks in the “shoe”, this strategy is called “basic” https://en.wikipedia.org/wiki/Blackjack#Basic_strategy
- The pre-hand expectation of the basic strategy is about -1% of the initial bet
- After the cards have been dealt, the CONDITIONAL expectation may be > 0
- The pre-hand CONDITIONAL expectation of “advantage” strategies can become slightly positive if you mentally keep track of the cards that have already been dealt from a finite shoe

Sampling without Replacement

- The hypergeometric distribution corresponds to sampling WITHOUT replacement and $\Pr(x|m, n, k) = \frac{\binom{m}{x} \binom{n}{k-x}}{\binom{m+n}{k}}$ where
 - m is the number of “good” elements in the set being drawn from
 - n is the number of “bad” elements in the set being drawn from
 - k is the number of times you draw without replacement
 - $\binom{a}{b} = \frac{a!}{b!(a-b)!}$ is the choose function, which is defined as zero if $a < b$
- What is the pre-hand probability of being dealt 21 in blackjack?

```
dhyper(x = 1, m = 4 * 4, n = 9 * 4, k = 1) * # probability of a 10, J, Q, or K
dhyper(x = 1, m = 4, n = 51 - 4, k = 1) * # probability of an Ace given a non-Ace already
2                                         # can get the Ace either first or second
```

```
## [1] 0.04826546
```

Blackjack Continued

- What is the probability of being dealt two cards of the same value in blackjack, in which case you have the option to split?

```
9 * dhyper(x = 2, m = 4, n = 12 * 4, k = 2) + # probability of a pair of A, 2, ..., 9  
    dhyper(x = 2, m = 4 * 4, n = 9 * 4, k = 2) # probability of two cards valued at 10
```

```
## [1] 0.1312217
```

- Suppose you have a 9 and a 2 from an infinite shoe. How would you calculate the probability that you end up winning if you double down against a dealer's 6?