More Probability with Discrete Random Variables

Ben Goodrich

January 29, 2019

Obligatory Disclosure

- Ben is an employee of Columbia University, which has received several research grants to develop Stan
- Ben is also a manager of GG Statistics LLC, which uses Stan for business purposes
- According to Columbia University policy, any such employee who has any equity stake in, a title (such as officer or director) with, or is expected to earn at least $\$5,000.00$ per year from a private company is required to disclose these facts in presentations

source("https://tinyurl.com/y93srfmp")

	0	1	2	3	4	5	6	7	8	9	10
0	0.000019	0.000019	0.000037	0.000056	0.000093	0.000149	0.000242	0.00039	0.000632	0.001022	0.001654
1	0.00003	0.00003	0.00006	0.00009	0.000151	0.000241	0.000392	0.000633	0.001025	0.001658	0
2	0.000098	0.000098	0.000196	0.000294	0.00049	0.000784	0.001274	0.002057	0.003331	0	0
3	0.000239	0.000239	0.000479	0.000718	0.001197	0.001916	0.003113	0.005029	0	0	0
4	0.000653	0.000653	0.001306	0.001959	0.003265	0.005225	0.00849	0	0	0	0
5	0.001724	0.001724	0.003448	0.005172	0.008621	0.013793	0	0	0	0	0
6	0.00467	0.00467	0.009339	0.014009	0.023348	0	0	0	0	0	0
7	0.012931	0.012931	0.025862	0.038793	0	0	0	0	0	0	0
8	0.036638	0.036638	0.073276	0	0	0	0	0	0	0	0
9	0.118534	0.118534	0	0	0	0	0	0	0	0	0
10	0.383621	0	0	0	0	0	0	0	0	0	0

Expectation of a Discrete Random Variable

```
round(Pr(Omega), digits = 3)  # What is the mode, median, and expectation of X1?
```

```
##     0     1     2     3     4     5     6     7     8     9    10
## 0.004 0.004 0.009 0.013 0.022 0.034 0.056 0.091 0.147 0.237 0.384
```

- 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 (9 here)
- EXPECTATION of a discrete random variable X is defined as

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

- Expectation is just a probability-weighted sum (8.431 here)

The Average Is an Estimator of an Expectation

- Since $\mu = \sum_{y \in \Omega} y \Pr\left(y\right)$, if we ESTIMATE $\Pr\left(y\right)$ with $\frac{1}{N} \sum_{n=1}^N \mathbb{I}\{y_n = y\}$,

$$\widehat{\mu} = \frac{1}{N} \sum_{y \in \Omega} y \sum_{n=1}^N \mathbb{I}\{y_n = y\} = \frac{1}{N} \sum_{y \in \Omega} \sum_{n=1}^N y \mathbb{I}\{y_n = y\} = \frac{1}{N} \sum_{n=1}^N y_n$$

- If we draw $\tilde{y}$ from its probability distribution S times, as $S \uparrow \infty$,

$$\frac{1}{S} \sum_{s=1}^S \tilde{y}_s \rightarrow \mu_Y$$

```
c(exact = sum(Omega * Pr(Omega)),
  approx = mean(sample(Omega, size = 10 ^ 8, replace = TRUE, prob = Pr(Omega))))
```

```
##    exact   approx
## 8.431034 8.431016
```

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 2 because the non-zero conditional probabilities are $\frac{1}{7}, \frac{1}{7}, \frac{2}{7}, \frac{3}{7}$
- How would we calculate the expectation of the second roll in a frame of bowling?

Marginal Expectation of Second Roll in Bowling

- To obtain $\mathbb{E}X_2$, we do

$$\begin{aligned}
\mathbb{E}X_2 &= \sum_{x_j \in \Omega_{X_2}} x_j \Pr\left(X_2 = x_j \middle| n = 10\right) \\
&= \sum_{x_j \in \Omega_{X_2}} x_j \sum_{x_i \in \Omega_{X_1}} \Pr\left(x_i \bigcap x_j\right) \\
&= \sum_{x_j \in \Omega_{X_2}} x_j \sum_{x_i \in \Omega_{X_1}} \Pr\left(x_j \middle| X_1 = x_i, n = 10\right) \Pr\left(x_i \middle| n = 10\right)
\end{aligned}$$

```
Pr_X2 <- colSums(joint_Pr) # marginal probabilies from last week
EX2 <- sum(Omega * Pr_X2)  # definition of marginal expectation
EX2
```

```
## [1] 1.064386
```

The Expectation Is a Linear Operator

- What is the expectation of cX where c is any constant?
- Answer: $c\mu$ because $\mathbb{E}\left[cX\right] = \sum_{x\in\Omega} cx\,\Pr\left(x\right) = c\sum_{x\in\Omega} x\,\Pr\left(x\right) = c\mathbb{E}X = c\mu$
- What is the expectation of the sum of two rolls in a frame of bowling?
- Answer: In general, $\mathbb{E}\left[aX + bY + c\right] = a\mu_X + b\mu_Y + c$, but in this case

$$\mathbb{E}\left[X+Y\right] = \sum_{x\in\Omega_X}\sum_{y\in\Omega_Y}(x+y)\Pr\left(x\bigcap y\right) = \sum_{x\in\Omega_X} x\overbrace{\sum_{y\in\Omega_Y}\Pr\left(x\bigcap y\right)}^{\Pr(x)} +$$

$$\sum_{y\in\Omega_Y} y\overbrace{\sum_{x\in\Omega_X}\Pr\left(x\bigcap y\right)}^{\Pr(y)} = \overbrace{\sum_{x\in\Omega_X} x\Pr\left(x\right)}^{\mathbb{E}[X]} + \overbrace{\sum_{y\in\Omega_Y} y\Pr\left(y\right)}^{\mathbb{E}[Y]} = \mu_X + \mu_Y$$

Sum of Two Rolls in a Frame

```
S <- row(joint_Pr) - 1 + col(joint_Pr) - 1; sum(joint_Pr * S)
```

```
## [1] 9.495421
```

	0	1	2	3	4	5	6	7	8	9	10
0	0	1	2	3	4	5	6	7	8	9	10
1	1	2	3	4	5	6	7	8	9	10	11
2	2	3	4	5	6	7	8	9	10	11	12
3	3	4	5	6	7	8	9	10	11	12	13
4	4	5	6	7	8	9	10	11	12	13	14
5	5	6	7	8	9	10	11	12	13	14	15
6	6	7	8	9	10	11	12	13	14	15	16
7	7	8	9	10	11	12	13	14	15	16	17
8	8	9	10	11	12	13	14	15	16	17	18
9	9	10	11	12	13	14	15	16	17	18	19
10	10	11	12	13	14	15	16	17	18	19	20

Expectations of Functions of Discrete RVs

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

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

- If $g\left(X\right)=\left(X-\mu\right)^{2}$, the VARIANCE of X is defined as $\mathbb{E}\left[\left(X-\mu\right)^{2}\right]=\sigma^{2}$. Show that $\sigma^{2}=\mathbb{E}\left[X^{2}\right]-\mu^{2}$ by expanding $\left(X-\mu\right)^{2}=X^{2}-2X\mu+\mu^{2}$.
- $\sigma=\sqrt[+]{\sigma^{2}}$ is the standard deviation but not an expectation of X
- If $g\left(X\right)=-\log(\Pr\left(X\right))$, the ENTROPY of X is $\mathbb{E}\left[-\log(\Pr\left(X\right))\right]$, which reaches its upper bound of `log(length(Omega))` when $\Pr\left(x\right)$ is constant

```
sum(-log(joint_Pr) * joint_Pr, na.rm = TRUE) # entropy
```

```
## [1] 2.361677
```

Expected Utility

- It is often sensible to make a decision that maximizes EXPECTED utility:

  1. Enumerate D possible decisions $\{d_1, d_2, \ldots, d_D\}$ that are under consideration
  2. Define a utility function $g\left(d, \ldots\right)$ that also depends on unknown (and maybe additional known) quantities
  3. Obtain / update your conditional probability distribution for all the unknowns given all the knowns
  4. Evaluate $\mathbb{E}g\left(d, \ldots\right)$ for each of the D decisions
  5. Choose the decision that has the highest value in (4)

- This is a very intuitive & useful procedure but you have to use Bayes Rule in (3)
- Also, whoever is deciding has to specify (1) and (2)

Iterated Expectations

- The expectation of a conditional expectation is a marginal expectation, i.e.

$$
\begin{aligned}
\mathbb{E}_X\left[\mathbb{E}\left[Y|\,X=x\right]\right] &= \mathbb{E}_X\left[\sum_{y\in\Omega_Y} y\,\mathrm{Pr}\left(y|\,X=x\right)\right] \\
&= \sum_{x\in\Omega_X}\mathrm{Pr}\left(X=x\right)\sum_{y\in\Omega_Y} y\,\mathrm{Pr}\left(y|\,X=x\right) \\
&= \sum_{x\in\Omega_Y}\sum_{y\in\Omega_X} y\,\mathrm{Pr}\left(y|\,X=x\right)\mathrm{Pr}\left(X=x\right) \\
&= \sum_{x\in\Omega_Y}\sum_{y\in\Omega_X} y\,\mathrm{Pr}\left(x\bigcap y\right) = \sum_{y\in\Omega_Y} y\sum_{x\in\Omega_X}\mathrm{Pr}\left(x\bigcap y\right) \\
&= \sum_{y\in\Omega_Y} y\,\mathrm{Pr}\left(y\right) = \mathbb{E}Y = \mu_Y
\end{aligned}
$$

Covariance and Correlation

- If $g\left(X,Y\right)=\left(X-\mu_X\right)\left(Y-\mu_Y\right)$, their COVARIANCE is defined as

$$\mathbb{E}g\left(X,Y\right)=\sum_{x\in\Omega_X}\sum_{y\in\Omega_Y}\left(x-\mu_X\right)\left(y-\mu_Y\right)\Pr\left(x\bigcap y\right)$$

- If $g\left(X,Y\right)=\frac{X-\mu_X}{\sigma_X}\times\frac{Y-\mu_Y}{\sigma_Y}$, their CORRELATION is defined as

$$\mathbb{E}g\left(X,Y\right)=\sum_{x\in\Omega_X}\sum_{y\in\Omega_Y}\frac{x-\mu_X}{\sigma_X}\frac{y-\mu_Y}{\sigma_Y}\Pr\left(x\bigcap y\right)=$$

$$\frac{1}{\sigma_X\sigma_Y}\sum_{x\in\Omega_X}\sum_{y\in\Omega_Y}\left(x-\mu_X\right)\left(y-\mu_Y\right)\Pr\left(x\bigcap y\right)=\frac{\mathrm{Cov}\left(X,Y\right)}{\sigma_X\sigma_Y}=\rho$$

- Covariance and correlation measure LINEAR dependence
- Is $\rho \substack{>\\=\\<} 0$ for 2 rolls in the same frame of bowling?

Correlation Calculation in R

```
Pr_X1 <- Pr(Omega)
EX1 <- sum(Omega * Pr_X1)
covariance <- 0
for (x1 in Omega) {
  for (x2 in 0:(10 - x1))
    covariance <- covariance + (x1 - EX1) * (x2 - EX2) * joint_Pr[x1 + 1, x2 + 1]
}
Var_X1 <- sum( (Omega - EX1) ^ 2 * Pr_X1 )
Var_X2 <- sum( (Omega - EX2) ^ 2 * Pr_X2 )
correlation <- covariance / sqrt(Var_X1 * Var_X2)
correlation
```

```
## [1] -0.8844158
```

Variance of a Sum

- What is the variance of the sum of two rolls in the same frame of bowling?

```
EX12 <- sum(joint_Pr * S)
Var_X12 <- sum( joint_Pr * (S - EX12) ^ 2 )
Var_X12
```

```
## [1] 0.8015148
```

- `Var_X12` is also equal to

```
Var_X1 + Var_X2 + 2 * covariance
```

```
## [1] 0.8015148
```

from the previous slide. How would you go about showing that is true in general?

Bernoulli Distribution

- $\Pr\left(X=1\middle|\, n=1\right) = \frac{\mathcal{F}_1}{-1+\mathcal{F}_{1+2}} = \frac{1}{-1+1-2} = \frac{1}{2}$ is one way to assign the probability of knocking down a single pin but is not the most general way
- The Bernoulli distribution over $\Omega = \{0,1\}$ is $\Pr\left(X=1\middle|\,\pi\right) = \pi \in [0,1]$ and thus $\Pr\left(X=0\middle|\,\pi\right) = 1-\pi$. Alternatively, $\Pr\left(x\middle|\,\pi\right) = \pi^x\left(1-\pi\right)^{1-x}$.
- What expression is the expectation of a Bernoulli random variable?
- What expression is the variance of a Bernoulli random variable?
- Why isn't the Bernoulli distribution appropriate for the pins in bowling?
- If $X_i = \mathbb{I}\{\text{pin i is knocked down}\}$ and π_i is the probability in the i-th Bernoulli distribution, what is

$$\Pr\left(x_1\middle|\,\pi_1\right)\prod_{i=2}^{10}\Pr\left(x_i\middle|\,\pi_i, X_1=x_1, X_2=x_2, \dots, X_{i-1}=x_{i-1}\right)?$$