GPCO 453: Quantitative Methods I

Sec 05: Probability, continued

Shane Xinyang Xuan[1]
ShaneXuan.com

October 29, 2017

[1]Department of Political Science, UC San Diego, 9500 Gilman Drive #0521.

Contact Information

Shane Xinyang Xuan

xxuan@ucsd.edu

The teaching staff is a team!

Professor Garg	Tu	1300-1500 (RBC 1303)
Shane Xuan	M	1100-1200 (SSB 332)
	M	1530-1630 (SSB 332)
Joanna Valle-luna	Tu	1700-1800 (RBC 3131)
	Th	1300-1400 (RBC 3131)
Daniel Rust	F	1100-1230 (RBC 3213)

Roadmap

In this section, we cover the basics for probability:

- Basic relationship of probability

Roadmap

In this section, we cover the basics for probability:

- Basic relationship of probability
- Independence and mutual exclusivity

Roadmap

In this section, we cover the basics for probability:

- Basic relationship of probability
- Independence and mutual exclusivity
- Bayes' Theorem

Roadmap

In this section, we cover the basics for probability:

- Basic relationship of probability
- Independence and mutual exclusivity
- Bayes' Theorem
- Z-score

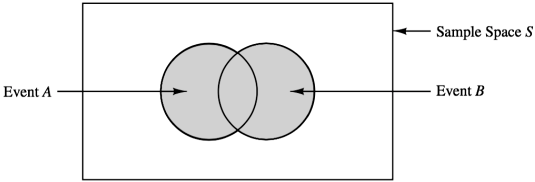
Roadmap

In this section, we cover the basics for probability:

- Basic relationship of probability
- Independence and mutual exclusivity
- Bayes' Theorem
- Z-score
- Probability distribution

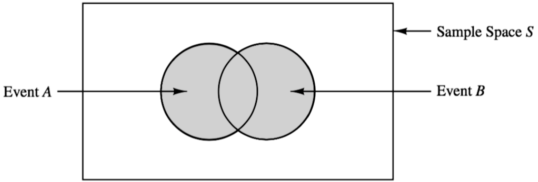
Basic Relationships of Probability

- Union (denoted by $\cup$): The union of A and B is the event containing **all** sample points belonging to A or B or both.

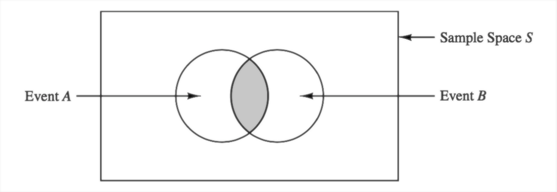


Basic Relationships of Probability

- Union (denoted by $\cup$): The union of A and B is the event containing **all** sample points belonging to A or B or both.

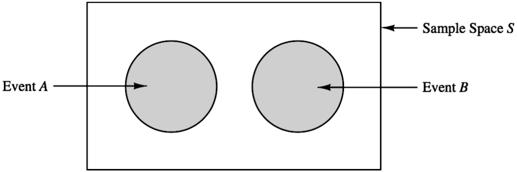


- Intersection (denoted by $\cap$): The intersection of A and B is the event containing the sample points belonging to **both** A and B.



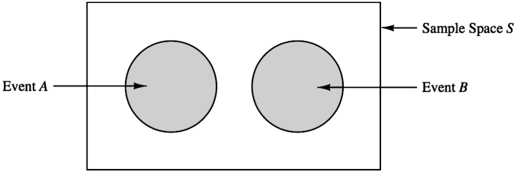
Independence and Mutual Exclusion

- Events A and B are mutually exclusive if, when one event occurs, the other **cannot** occur; $\Pr(A \cup B) = \Pr(A) + \Pr(B)$



Independence and Mutual Exclusion

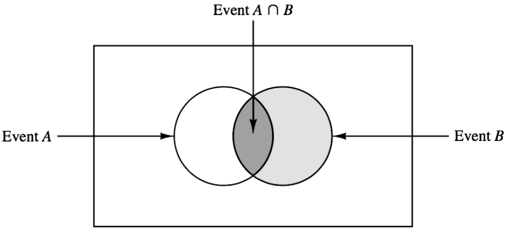
- Events A and B are mutually exclusive if, when one event occurs, the other **cannot** occur; $\Pr(A \cup B) = \Pr(A) + \Pr(B)$



- Two events A and B are independent if

$$\Pr(A|B) = \Pr(A) \quad (1)$$
$$\Pr(B|A) = \Pr(B) \quad (2)$$

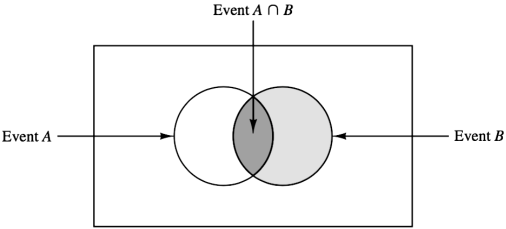
More on Conditional Probability



$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)} \qquad \text{conditional probability} \qquad (3)$$

$$\Pr(B|A) = \frac{\Pr(A \cap B)}{\Pr(A)} \qquad \text{conditional probability} \qquad (4)$$

More on Conditional Probability



$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)} \quad \text{conditional probability} \quad (3)$$

$$\Pr(B|A) = \frac{\Pr(A \cap B)}{\Pr(A)} \quad \text{conditional probability} \quad (4)$$

- Multiplication law

$$\Pr(A \cap B) = \Pr(B)\Pr(A|B) \quad \text{general case} \quad (5)$$

$$\Pr(A \cap B) = \Pr(B)\Pr(A) \quad \text{independent events} \quad (6)$$

Bayes' Theorem (Two-Event Case)

$$\Pr(A_1|B) = \frac{\Pr(A_1)\Pr(B|A_1)}{\Pr(A_1)\Pr(B|A_1) + \Pr(A_2)\Pr(B|A_2)} \tag{7}$$

$$\Pr(A_2|B) = \frac{\Pr(A_2)\Pr(B|A_2)}{\Pr(A_1)\Pr(B|A_1) + \Pr(A_2)\Pr(B|A_2)} \tag{8}$$

Bayes' Theorem (Two-Event Case)

$$\Pr(A_1|B) = \frac{\Pr(A_1)\Pr(B|A_1)}{\Pr(A_1)\Pr(B|A_1) + \Pr(A_2)\Pr(B|A_2)} \tag{7}$$

$$\Pr(A_2|B) = \frac{\Pr(A_2)\Pr(B|A_2)}{\Pr(A_1)\Pr(B|A_1) + \Pr(A_2)\Pr(B|A_2)} \tag{8}$$

- Important to note that

$$\Pr(A_1) + \Pr(A_2) = 1 \tag{9}$$

$$\Pr(A_1 \cap B) + \Pr(A_2 \cap B) = \Pr(B) \tag{10}$$

Example: Bayes' Theorem

- 983,764 putts were made and 629,470 putts were missed.

Example: Bayes' Theorem

- 983,764 putts were made and 629,470 putts were missed.
  $\Pr(\mathsf{putt}) = \frac{983764}{1613234} \approx .61$; $\Pr(\neg\mathsf{putt}) = 1 - .61 = .39$

Example: Bayes' Theorem

- 983,764 putts were made and 629,470 putts were missed.
  $\Pr(\text{putt}) = \frac{983764}{1613234} \approx .61$; $\Pr(\neg\text{putt}) = 1 - .61 = .39$
- Suppose that a PGA Tour player has a par putt. It is known that of putts made, 64.0% were for par whereas for putts missed, 20.3% were for par. What is the revised probability of making a putt given the PGA Tour player has a par putt?

Example: Bayes' Theorem

- 983,764 putts were made and 629,470 putts were missed.
  $\Pr(\text{putt}) = \frac{983764}{1613234} \approx .61$; $\Pr(\neg\text{putt}) = 1 - .61 = .39$
- Suppose that a PGA Tour player has a par putt. It is known that of putts made, 64.0% were for par whereas for putts missed, 20.3% were for par. What is the revised probability of making a putt given the PGA Tour player has a par putt?

$$\Pr(\text{putt}|\text{par}) = \frac{\Pr(\text{putt})\Pr(\text{par}|\text{putt})}{\Pr(\text{putt})\Pr(\text{par}|\text{putt}) + \Pr(\neg\text{putt})\Pr(\text{par}|\neg\text{putt})}$$
$$= \frac{.61 \times .64}{.61 \times .64 + .39 \times .203} \approx .831 \qquad (11)$$

Z-score

- Sometimes we want to standardize a random variable

$$Z \equiv \frac{X - \mu_X}{\sigma_X}, \tag{12}$$

where $\sigma_X = \sqrt{\sigma_X^2}$.

Z-score

- Sometimes we want to standardize a random variable

$$Z \equiv \frac{X - \mu_X}{\sigma_X}, \tag{12}$$

where $\sigma_X = \sqrt{\sigma_X^2}$.

- Motivation of standardizing a random variable:

$$\begin{aligned} \text{Var}(Z) &= \frac{1}{\sigma_X^2}\text{Var}(X) && (13)\\ &= \frac{1}{\sigma_X^2}\sigma_X^2 && (14)\\ &= 1 && (15) \end{aligned}$$

Z-score

- Sometimes we want to standardize a random variable

$$Z \equiv \frac{X - \mu_X}{\sigma_X}, \tag{12}$$

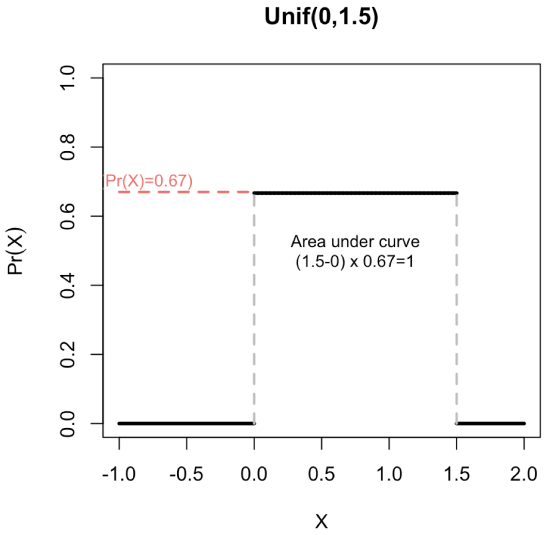
where $\sigma_X = \sqrt{\sigma_X^2}$.

- Motivation of standardizing a random variable:

$$\begin{aligned} \text{Var}(Z) &= \frac{1}{\sigma_X^2}\text{Var}(X) && (13) \\ &= \frac{1}{\sigma_X^2}\sigma_X^2 && (14) \\ &= 1 && (15) \end{aligned}$$

- **Outliers** are defined as observations having a Z-score below -3 or more than 3.

Uniform Distribution



Uniform Distribution

- The density function is

$$f(x) = \frac{1}{b - a}, a \leq x \leq b \tag{16}$$

Uniform Distribution

- The density function is

$$f(x) = \frac{1}{b - a}, a \leq x \leq b \tag{16}$$

- $E[x] = \frac{a+b}{2}$

Uniform Distribution

- The density function is

$$f(x) = \frac{1}{b - a}, \ a \le x \le b \tag{16}$$

- $E[x] = \frac{a+b}{2}$
- $\text{var}(x) = \frac{(b-a)^2}{12}$

Uniform Distribution

- The density function is

$$f(x) = \frac{1}{b - a}, \; a \leq x \leq b \tag{16}$$

- $E[x] = \frac{a+b}{2}$
- $\text{var}(x) = \frac{(b-a)^2}{12}$
- The expected time for a plane to land is 30 minutes. What is the probability of landing between 25 to 30 minutes?

Uniform Distribution

- The density function is

$$f(x) = \frac{1}{b - a}, \; a \leq x \leq b \tag{16}$$

- $E[x] = \frac{a+b}{2}$
- $\text{var}(x) = \frac{(b-a)^2}{12}$
- The expected time for a plane to land is 30 minutes. What is the probability of landing between 25 to 30 minutes?
- Density of probability is $\frac{1}{30-0} = \frac{1}{30}$

Uniform Distribution

- The density function is

$$f(x) = \frac{1}{b - a}, \; a \leq x \leq b \tag{16}$$

- $E[x] = \frac{a+b}{2}$
- var$(x) = \frac{(b-a)^2}{12}$
- The expected time for a plane to land is 30 minutes. What is the probability of landing between 25 to 30 minutes?
- Density of probability is $\frac{1}{30-0} = \frac{1}{30}$
- Interval of success is $[25, 30]$

Uniform Distribution

- The density function is

$$f(x) = \frac{1}{b - a}, \ a \leq x \leq b \tag{16}$$

- $E[x] = \frac{a+b}{2}$
- $\text{var}(x) = \frac{(b-a)^2}{12}$
- The expected time for a plane to land is 30 minutes. What is the probability of landing between 25 to 30 minutes?
- Density of probability is $\frac{1}{30-0} = \frac{1}{30}$
- Interval of success is $[25, 30]$
- Area under the curve is $\underbrace{(5)}_{\text{length}} \underbrace{\left(\frac{1}{30}\right)}_{\text{height}} = \frac{1}{6}$

Binomial Distribution

- The probability of having x successes in n trials is

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x} \tag{17}$$

Binomial Distribution

- The probability of having x successes in n trials is

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x} \tag{17}$$

  - x: $\#$(success)

Binomial Distribution

- The probability of having x successes in n trials is

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x} \tag{17}$$

  - x: #(success)
  - p: probability of success

Binomial Distribution

- The probability of having x successes in n trials is

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x} \tag{17}$$

  - x: #(success)
  - p: probability of success
  - n: #(trials)

Binomial Distribution

- The probability of having x successes in n trials is

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x} \tag{17}$$

  - x: #(success)
  - p: probability of success
  - n: #(trials)
- $\mathbb{E}[x] = np$

Binomial Distribution

- The probability of having x successes in n trials is

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x} \tag{17}$$

  - x: $\#$(success)
  - p: probability of success
  - n: $\#$(trials)
- $\mathbb{E}[x] = np$
- $\text{var}(x) = np(1-p)$

Binomial Distribution: Example

- A university found that 20% of its students withdraw without completing the introductory statistics course. Assume that 20 students registered for the course. Compute the probability that more than 2 will withdraw.

$$
\begin{aligned}
\Pr(X > 2) &= 1 - \Pr(X = 0) - \Pr(X = 1) - \Pr(X = 2) \\
&= 1 - \binom{20}{0} .2^0 (1 - .2)^{20-0} - \binom{20}{1} .2^1 (.8)^{20-1} - \binom{20}{2} .2^2 (.8)^{20-2}
\end{aligned}
$$

Binomial Distribution: Example

- A university found that 20% of its students withdraw without completing the introductory statistics course. Assume that 20 students registered for the course. Compute the probability that more than 2 will withdraw.

$$
\begin{aligned}
\Pr(X > 2) &= 1 - \Pr(X = 0) - \Pr(X = 1) - \Pr(X = 2) \\
&= 1 - \binom{20}{0} .2^0 (1 - .2)^{20-0} - \binom{20}{1} .2^1 (.8)^{20-1} - \binom{20}{2} .2^2 (.8)^{20-2}
\end{aligned}
$$

- Compute the expected number of withdrawals.

$$\mathbb{E}[X] = np = 20 \times .2 = 4 \tag{18}$$

Poisson Distribution

- The probability of x occurrences in an interval is

$$f(x) = \frac{\mu^x e^{-\mu}}{x!} \tag{19}$$

Poisson Distribution

- The probability of x occurrences in an interval is

$$f(x) = \frac{\mu^x e^{-\mu}}{x!} \tag{19}$$

  - μ is the expected value of occurrences in an interval

Poisson Distribution

- The probability of x occurrences in an interval is

$$f(x) = \frac{\mu^x e^{-\mu}}{x!} \tag{19}$$

  - μ is the expected value of occurrences in an interval
  - x: #(occurrence)

Poisson Distribution

- The probability of x occurrences in an interval is

$$f(x) = \frac{\mu^x e^{-\mu}}{x!} \tag{19}$$

  - μ is the expected value of occurrences in an interval
  - x: #(occurrence)
- Both mean and variance are μ

Poisson Distribution

- The probability of x occurrences in an interval is

$$f(x) = \frac{\mu^x e^{-\mu}}{x!} \tag{19}$$

  - μ is the expected value of occurrences in an interval
  - x: #(occurrence)

- Both mean and variance are μ
- During the period of time that a local university takes phone-in registrations, calls come in at the rate of one every two minutes. What is the probability of three calls in five minutes?

$$\Pr(\text{three calls in five minutes}) = \frac{(2.5)^3 e^{-2.5}}{3!}$$

Appendix: Code for Uniform Distribution

```
x <- seq(0,1.5,length=5000)
unif <- dunif(x, min=0, max=1.5)
plot(y=unif, x=x, type="l", lwd=3, xlim=c(-1,2), ylim=c(0,1),
     main="Unif(0,1.5)", ylab=expression(Pr(X)), xlab='X')
segments(x0=-1, x1=0, y0=0, y1=0, lwd=3)
segments(x0=1.5, x1=2, y0=0, y1=0, lwd=3)
segments(x0=0, x1=0, y0=0, y1=0.67, lty=2, lwd=2, col="gray")
segments(x0=1.5, x1=1.5, y0=0, y1=0.67, lty=2, lwd=2, col="gray"
segments(x0=-1, x1=0, y0=0.67, y1=0.67, lty=2, lwd=2, col="india
text(-.8,.7,"(Pr(X)=0.67)", col="indianred1", cex=.85)
text(.75,.5,"Area under curve\n(1.5-0) x 0.67=1", cex=.85)
```