Lecture 11
Probability

Ryan McWay[†]

[†] *Applied Economics,*
University of Minnesota

Mathematics Review Course, Summer 2023
University of Minnesota
August 21st, 2023

Last Lecture Review

- (Inequality) Constrained Optimization:
  - Kuhn Tucker Conditions
  - Corner Solutions
- Comparative Statics & Envelope Theorem:
  - The Multiplier
  - Comparative Statics
  - Envelope Theorem

Review Assignment

1. Problem Set 10 solutions are available on Github.
2. Any issues or problems **You** would like to discuss?

Daily Icebreaker

- Attendance via prompt:
  - Name
  - Daily Icebreaker: What crime would you commit if you would either never be caught for it or had complete immunity from conviction?

Probability

Motivation

- General background
  - How we evaluate the likelihood of events occurring.
- Why do economists' care?
  - This provides a foundation for 'empirical' economics to work with real-world data.
- Application in this career
  - In working with estimators and throughout econometrics.

Motivation

- General background
  - How we evaluate the likelihood of events occurring.
- Why do economists' care?
  - This provides a foundation for 'empirical' economics to work with real-world data.
- Application in this career
  - In working with estimators and throughout econometrics.

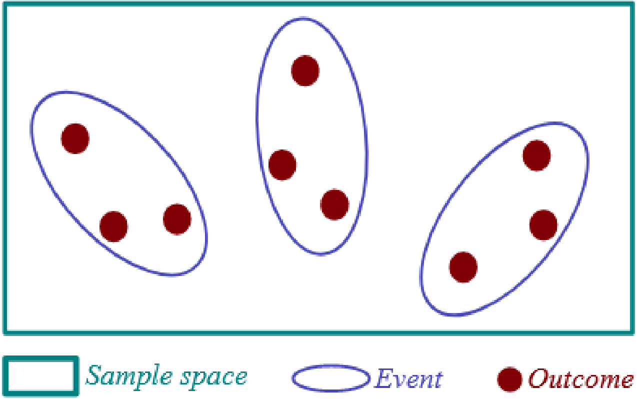
Motivation

- General background
  - How we evaluate the likelihood of events occurring.
- Why do economists' care?
  - This provides a foundation for 'empirical' economics to work with real-world data.
- Application in this career
  - In working with estimators and throughout econometrics.

Overview

1. Outcomes and Events

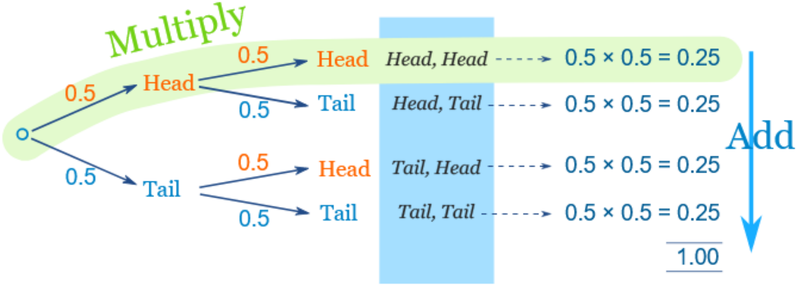
- Outcomes: All possible values that may be realized given the domain.
- Sample space: The set of all possible outcomes.
- Event: A subset of the outcomes in the sample space.



2. Probability

- Experiment: A procedure that could be infinitely repeated with a well-defined set of outcomes.
- Random Trail: One run of the experiment.
- Relative Frequency: Fraction of random trails in which an event occurs.
- Probability $Pr(A)$: The relative frequency approached in the limit as the experiment is repeated infinitely.
  - How likely an outcome will occur in any given random trail.
- Probability Tree: A diagram of potential outcomes determining the probability of occurrence.
- Complement A^c: All other events except those that occur in event A.
  - $Pr(A^c) = 1 - Pr(A)$

Probability Tree



2. Probability

- Probabilities are ranged $0-1$, with the sum of all possible events equaling 1.
  - $Pr(A) \leq 1$
  - $\sum_i^n Pr(A_i) = 1$
- Non-existent Event: $Pr(\emptyset) = 0$
- Monotone Probability Inequality: If $A \subset B \implies Pr(A) \leq Pr(B)$
- Inclusion-Exclusion Principle: $Pr(A \cup B) = Pr(A) + Pr(B) - Pr(A \cap B)$
- Boole's Inequality: $Pr(A \cup B) \leq Pr(A) + Pr(B)$
- Bonferroni's Inequality: $Pr(A \cap B) \geq Pr(A) + Pr(B) - 1$

3. Probability Limits

- Probability limits are used to test large sample properties of a distribution or estimator.
- This relies on asymptotics of the distribution as $n \rightarrow \infty$.
- Probability limit approaching c:

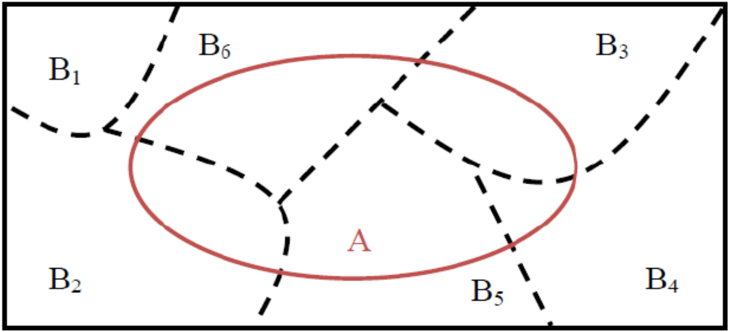
$$\begin{aligned} \lim_{n\rightarrow\infty} Pr([x_n - c] > \varepsilon) =& 0 \\ x_n =& c \\ x_n \xrightarrow{p} & c \end{aligned}$$

4. Independence

- Independence: Events where the probability of an event is unrelated to the outcome of another event.
- $Pr(A|B) = Pr(A)$
- $\implies Pr(A \wedge B) = Pr(A|B) * Pr(B) = Pr(A) * Pr(B)$
- Mutually Exclusive: Each outcome have nothing in common (do not share the sample space).
- $E_i \cap E_j = \emptyset$

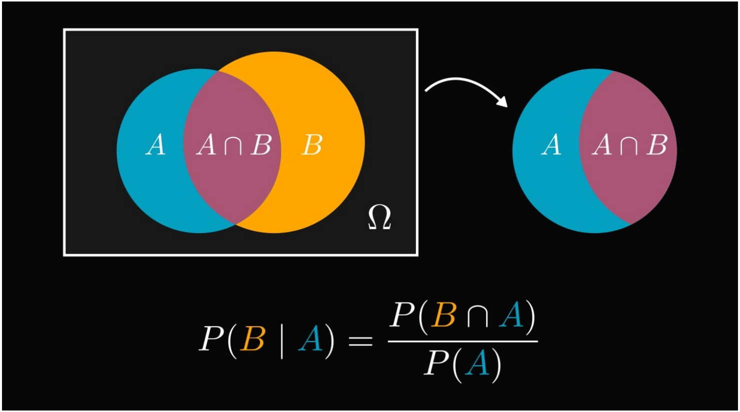
5. Law of Total Probability

- Suppose the sample space is partitioned into n disjoint events: $B_1, \ldots, B_n$.
- $Pr(A) = \sum_i^n Pr(A|B_i) * Pr(B_i)$



6. Conditional Probability

- $Pr(A|B)$: The probability of A occurring given B occurs.
- $Pr(A|B) = Pr(A \wedge B)Pr(B)$



Demonstration: Conditional Probability

Question:

The probability a product breaks down is $P(T \geq t) = e^{-t/5} \forall t \geq 0$. I have the product for two years and didn't breakdown. What is the probability it breaks down in year 3?

Answer:

$$
\begin{aligned}
P(B) =& P(T \geq 2) = e^{-2/5} \\
P(A) = P(A \cap B) =& P(2 \leq T \leq 3) = e^{-2/5} - e^{-3/5} \\
P(A|B) =& \frac{P(A \cap B)}{P(B)} \\
=& \frac{e^{-2/5} - e^{-3/5}}{e^{-2/5}} \approx 0.1813
\end{aligned}
$$

Demonstration: Conditional Probability

Question:

The probability a product breaks down is $P(T \geq t) = e^{-t/5} \forall t \geq 0$. I have the product for two years and didn't breakdown. What is the probability it breaks down in year 3?

Answer:

$$
\begin{aligned}
P(B) =& P(T \geq 2) = e^{-2/5} \\
P(A) = P(A \cap B) =& P(2 \leq T \leq 3) = e^{-2/5} - e^{-3/5} \\
P(A|B) =& \frac{P(A \cap B)}{P(B)} \\
=& \frac{e^{-2/5} - e^{-3/5}}{e^{-2/5}} \approx 0.1813
\end{aligned}
$$

7. Cumulative Distribution Function

- $F(x)$: Describes the probability that a random variable x is less than or equal to that particular realization.
- $0 \leq F(X) \leq 1$
- Discrete:

$$F(x_j) = Pr(X \leq x_j)$$

- Continuous:

$$F(x_j) = \int_{-\infty}^{x_j} f(x)dx$$

7. Cumulative Distribution Function

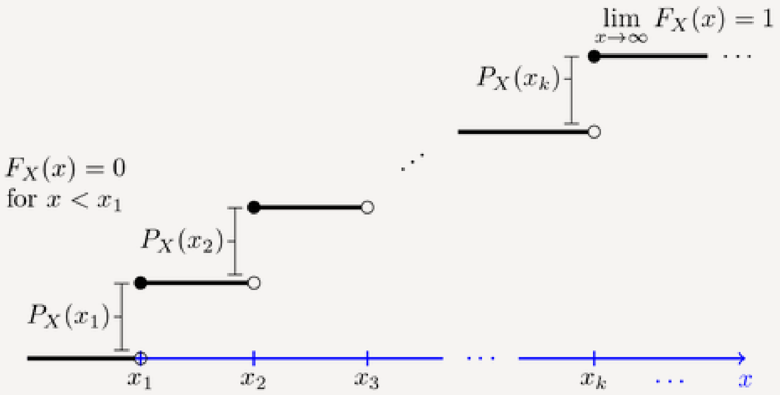
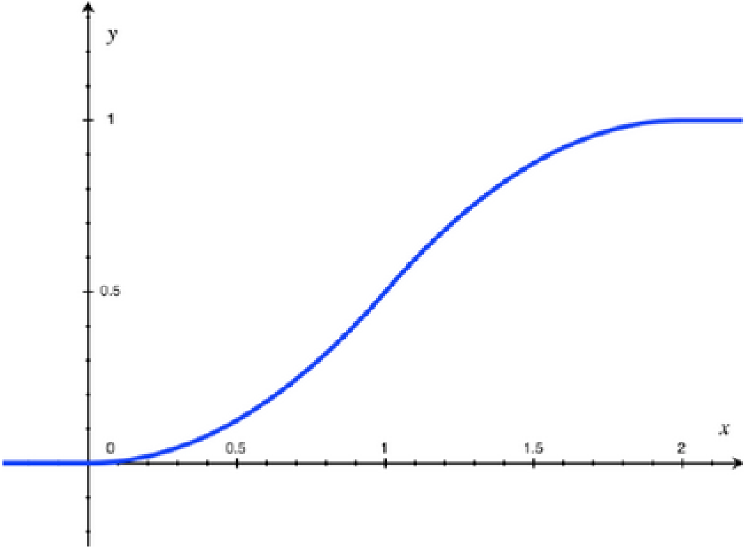


Fig.3.4 - CDF of a discrete random variable.

7. Cumulative Distribution Function



8. Probability Distribution Function

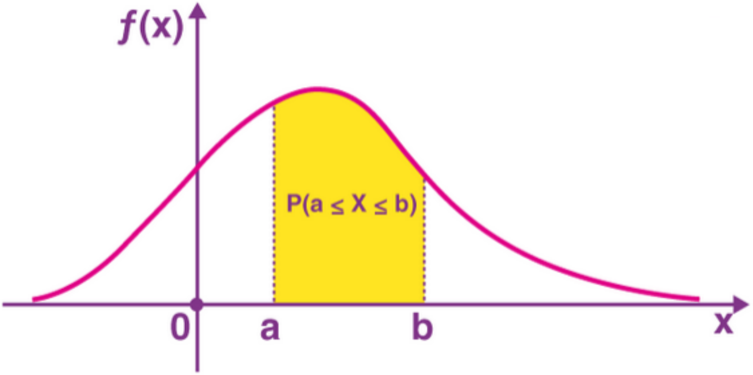
- $f(x)$: The frequency of a random variable outcomes across the sample space.
- $0 \leq f(x) \leq 1$
- Discrete:

$$f(x_j) = \begin{cases} p_j & \text{where } j = 1, \ldots, k \\ 0 & \text{otherwise} \end{cases}$$

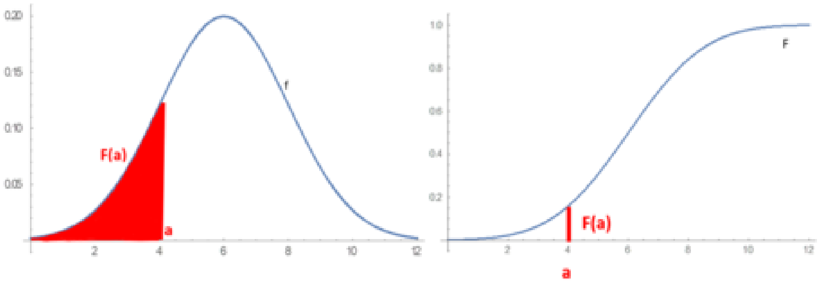
- Continuous:

$$Pr(a \leq X \leq b) = \int_a^b f(x)dx$$

8. Probability Distribution Function



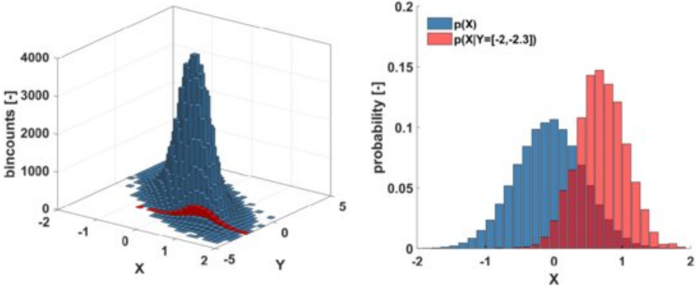
PDF TO CDF



9. Conditional Probability Distribution Function

- Note that $Pr(A|B) = \frac{Pr(A \wedge B)}{Pr(B)}$.
- Implies $Pr(A \wedge B) = Pr(A|B) \times Pr(B)$.
- Using this logic: $f_{Y|X}(y|x) = f(y|x) = Pr(Y = y|X = x)$.
- Implies $f(y|x) = \frac{f(x,y)}{f(x)}$
- If X and Y are independent, then $f(y|x) = f(y) \forall x, y$.
- Implies $f_{X,Y}(x, y) = f_X(x) \times f_Y(y) \forall x, y$

9. Conditional Probability Distribution Function



10. Joint and Marginal Distributions

- Joint Distribution $Pr(A \wedge B)$: The joint probability density functions.
- Discrete:

$$f_{X,Y}(x,y) = Pr(X = x \wedge Y = y)$$

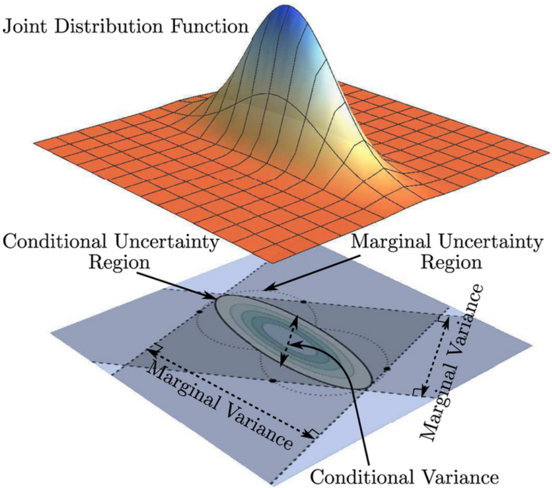
- Continuous:

$$Pr(a \leq x \leq b \wedge c \leq y \leq d) = \int_c^d \int_a^b f(x,y)dxdy$$

- Marginal Distribution $f(x)$: The joint distribution for x conditioning on all values of y.

$$f(x) = \sum_y f(x,y)$$

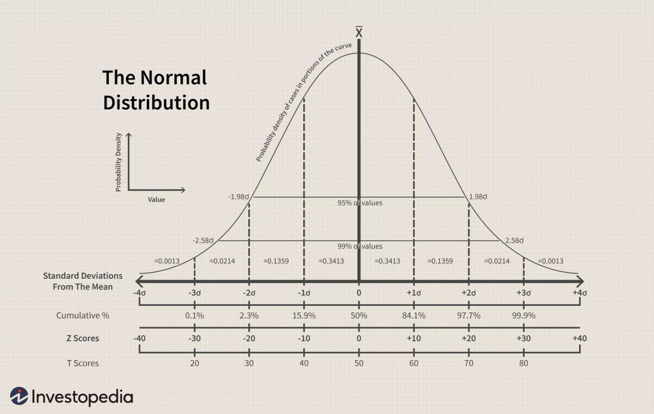
10. Joint and Marginal Distributions



11. Gaussian (Normal) Distribution

$$f(x|\mu,\sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{(x-\mu)^2}{\sigma^2}\right)}$$

$$X \sim N(\mu, \sigma^2)$$



12. Other Distributions

- Chi-Squared Distribution

$$X = \sum_{i=1}^{n} Z_i^2 \sim \mathcal{X}_n^2$$

- T Distribution

$$T = \frac{Z}{\sqrt{\frac{x}{n}}} \sim t_n$$

- F Distribution

$$F = \frac{X_1/n_1}{X_2/n_2} \sim F_{n_1,n_2}$$

- Logistical Distribution

$$f(x) = \frac{1}{1+e^{-x}}(1 - \frac{1}{1+x^{-x}})$$

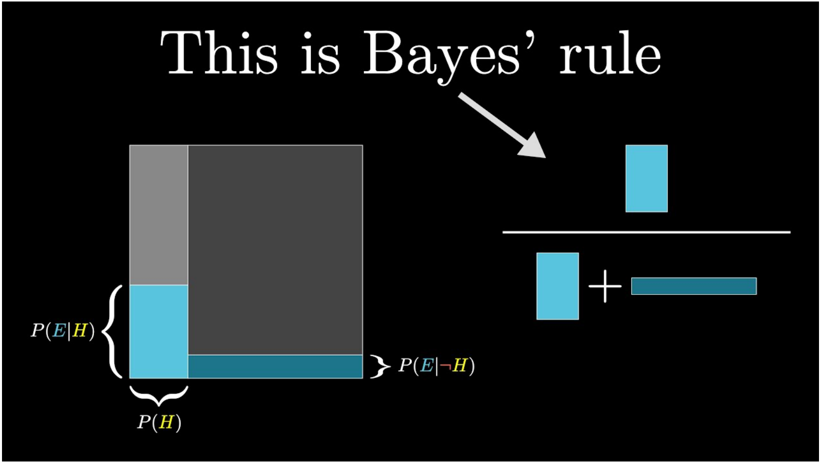
13. Bayes Rule

- As compared to a frequentest approach (what we have been discussing), Bayesian probabilities condition the frequency of an event occurring on the prior information known about the other events occurring.
- For two events A and B:

$$Pr(B|A) = \frac{Pr(B) \times Pr(A|B)}{Pr(A)}$$

$$Pr(B|A) = \frac{Pr(B) \times Pr(A|B)}{Pr(A|B) \times Pr(B) + Pr(A|B^c) \times Pr(B^c)}$$

13. Bayes Rule



Demonstration: Bayes Rule

Question:

A patient's probability of having a disease is $P(A) = 0.1$. The patient's probability of having an underlying condition is $P(B) = 0.05$. Being informed, you know $P(B|A) = 0.07$ is the probability the patient has the condition given the disease. What is the risk a patient has the disease given they have the condition $P(A|B)$?

Answer:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \frac{0.07 \times 0.1}{0.05} = 0.14$$

Demonstration: Bayes Rule

Question:

A patient's probability of having a disease is $P(A) = 0.1$. The patient's probability of having an underlying condition is $P(B) = 0.05$. Being informed, you know $P(B|A) = 0.07$ is the probability the patient has the condition given the disease. What is the risk a patient has the disease given they have the condition $P(A|B)$?

Answer:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \frac{0.07 \times 0.1}{0.05} = 0.14$$

Practice: Bayes Rule

1. 1% of women over 50 have breast cancer. 90% of women with breast cancer had a positive test (e.g., mammogram). 8% of all women get a false positive test (i.e., positive when there is **no** cancer). What is the probability of a woman having cancer if she has a positive mammogram test?

PRACTICE: BAYES RULE

1. 1% of women over 50 have breast cancer. 90% of women with breast cancer had a positive test (e.g., mammogram). 8% of all women get a false positive test (i.e., positive when there is **no** cancer). What is the probability of a woman having cancer if she has a positive mammogram test?

Answer: Show Work

$$Pr(A|B) \approx 8.65\%$$

Practice: Bayes Rule

1. Female cancer rates.
2. 50% of email is spam. A spam filter can detect 99% of spam. The probability of a false positive (i.e., not spam) is 5%. What is the probability an email in your spam folder is **not** spam?

Practice: Bayes Rule

1. Female cancer rates.
2. 50% of email is spam. A spam filter can detect 99% of spam. The probability of a false positive (i.e., not spam) is 5%. What is the probability an email in your spam folder is **not** spam?

Answer: Show Work

$$P(B^c|A) = \frac{5}{104}$$

14. Moments of a Distribution

- 1st Moment: Mean or average occurrence.

$$\bar{X} = \mu_x = \frac{1}{n}\sum_{i=1}^{n} X_i$$

- 2nd Moment: Variance or spread of distribution

$$Var(X) = \sigma_x = \mathbb{E}[(X - \mu_x)^2]$$

- 3rd Moment: Skewness or lack of symmetry in the distribution

$$Skew(X) = \frac{\mu_x^3}{\sigma_x^3} = \mathbb{E}[(X - \mu_x)^3]$$

- 4th Moment: Kurtosis or relative weight (fatness) of the tails of the distribution

$$Kurt(X) = \frac{\mu_x^4}{\sigma_x^4} - 3 = \mathbb{E}[(X - \mu_x)^4]$$

15. Variance & Standard Deviation

- Variance: The spread of the distribution. How typical is a value x given the mean $\bar{X}$.

$$Var(X) = \sigma_x^2 = \mathbb{E}[(X - \mu_x)^2]$$

- If the distribution is independently and identically distributed (I.I.D.)

$$Var(X) = \mathbb{E}[(X - \mu_x)^2] = \mathbb{E}[(X^2)] - (\mathbb{E}[X])^2$$

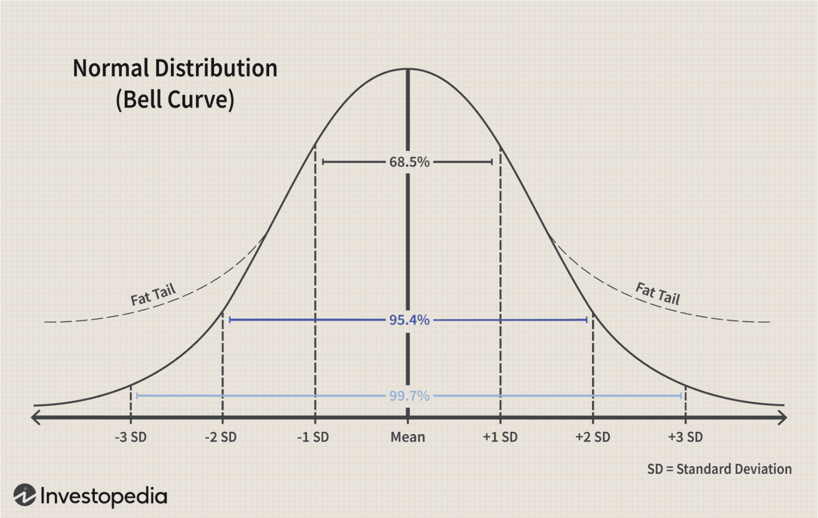
- In a sample, variance is adjusted by sample size:

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

- Standard Deviation: The typical deviation from the mean.

$$sd(X) = \sigma_x = +\sqrt{Var(X)} \equiv +\sqrt{s^2}$$

15. Variance & Standard Deviation



16. Covariance

- Indicator for degree two variables move together (co-vary) as either variable moves about its own distribution.

$$Cov(X, Y) = \sigma_{XY} = \mathbb{E}[(X - \mu_X)(Y - \mu_Y)]$$

17. Correlation

- Indicator for degree one variables move due to changes in another variable (e.g., Not their combined variance)
- $-1 \leq Corr(X, Y) \leq 1$
- Independence: $Corr(X, Y) = 0$

$$Corr(X, Y) = \frac{Cov(X, Y)}{sd(X) \times sd(Y)} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

Review

Review: Probability

1. Outcomes & Events
2. Probability
3. Probability Limits
4. Independence
5. Law of Total Probability
6. Conditional Probability
7. Cumulative Distribution Function
8. Probability Distribution Function
9. Conditional Probability Distribution Function
10. Joint & Marginal Distributions
11. Gaussian (Normal) Distribution
12. Other Distributions
13. Bayes Rules
14. Moments of a Distribution
15. Variance & Standard Deviation
16. Covariance
17. Correlation

Assignment

- Readings on Statistics before Lecture 12:
  - Hansen Metrics Ch. 6, 7, & 8
- Assignment:
  - Problem Set 11 (PS11)
  - Solution set will be available following end of Lecture 11
- Struggling?
  1. Read the ‘Encouraged Reading’
  2. Review ‘Supplementary material’
  3. Reach out directly

Bayes Rule Question 1 Answer:

◂ Question

$$
\begin{aligned}
Pr(B) &= 0.01 \\
Pr(B^c) &= 0.99 \\
Pr(A|B) &= 0.9 \\
Pr(A|B^c) &= 0.096 \\
Pr(B|A) &= \frac{0.9 \cdot 0.01}{0.9 \cdot 0.01 + 0.096 \cdot 0.99}
\end{aligned}
$$

Bayes Rule Question 2 Answer:

Question

A = spam detected. B = email is spam. B^c = email is not spam.

$$
\begin{aligned}
Pr(B) = P(B^c) &= 0.5 \\
Pr(A|B) &= 0.99 \\
Pr(A|B^c) &= 0.05 \\
Pr(B^c|A) &= \frac{Pr(A|B^c)Pr(B^c)}{Pr(A|B)P(B) + Pr(A|B^c)Pr(B^c)} \\
&= \frac{0.05 \cdot 0.5}{0.99 \cdot 0.5 + 0.05 \cdot 0.5}
\end{aligned}
$$