All of Statistics: A Concise Course in Statistical Inference

By Larry Wasserman

Chapter 1 & 2 Overview Presentation

Toby Xu
UW Madison
05/29/07

Chapter 1 Overview

- Sample Space
- Disjoint or Mutually Exclusive
- Probability Axioms/Basic Properties
  - Finite sample spaces
  - Independent Events
  - Conditional Probability
    - Baye’s Theorem

Sample Space

- **Sample Space** Ω is the set of possible outcomes.
- Ex: Toss fair coin twice Ω={HH,HT,TH,TT}
- **Points** ω in Ω are called **sample outcomes, realizations, elements.**
- **Subsets** of Ω are called **events.**
- Ex: Event that first toss is head is A={HH,HT}
- A^c = complement of A (or NOT A)
- $\varnothing$ = complement of Ω

$$A \bigcup B = \{\omega \in \Omega : \omega \in A \ or \ \omega \in B \ or \ \omega \in both\}$$

$$A \bigcap B = \{\omega \in \Omega : \omega \in A \ and \ \omega \in B\}$$

Disjoint or Mutually Exclusive

- Two events (A and B) are mutually exclusive iff $A \bigcap B = \varnothing$

  Ex: A_1=[0,1), A_2=[1,2), A_3=[2,3)

  - A Partition of Ω is a sequence of disjoint sets
  - Indicator function of A
    - Monotone increasing if $A_1 \subset A_2 \subset A_3 \cdots$
    - Monotone decreasing if $A_1 \supset A_2 \supset A_3 \cdots$

Intro to Probability

- P(A) = Probability Distribution or Probability Measure
- Axioms:
  - 1) $P(A) \geq 0$ for every A
  - 2) $P(\Omega) = 1$
  - If $A_1, A_2 \ldots$ are disjoint then

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

- Statistical Interference:
  - Frequentist
  - Bayesian Schools

Basic Properties of Probability

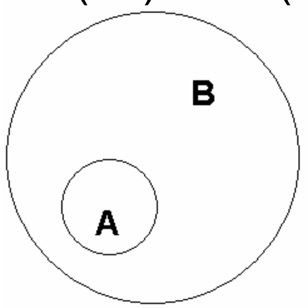
$$P(\varnothing) = 0$$

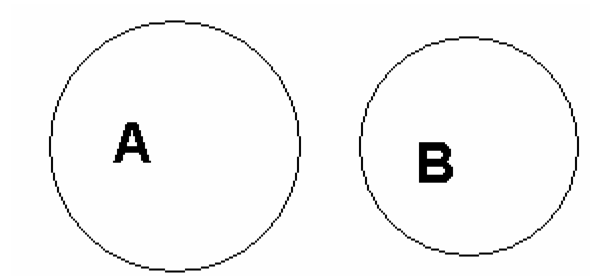
$$A \subset B \Rightarrow P(A) \le P(B)$$

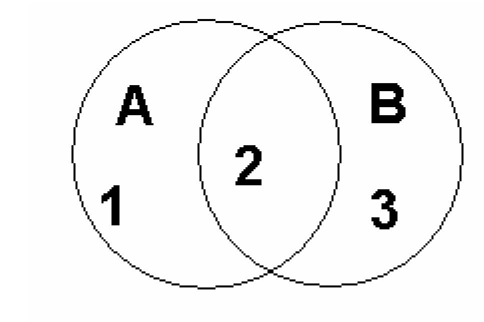
$$0 \le P(A) \le 1$$

$$A \bigcap B = \varnothing \Rightarrow P(A \bigcup B) = P(A) + P(B)$$ For Disjoint Probabilities only

$P(A^c)=1-P(A)$







$$P(A \bigcup B) = P(A) + P(B) - P(AB)$$

A or B = (AB^c) or (AB) or (A^cB)

Probability on Finite Sample Spaces

- P(A)=A/Ω

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

N choose k

$$\binom{n}{0} = \binom{n}{n} = 1$$

$$\binom{n}{k} = \binom{n}{n-k}$$

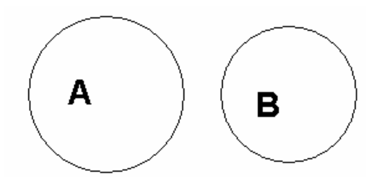
- N choose K is counting how many ways can we get a *k*-subset out from a set with *n* elements.

- Ex: There's 10 people in the Book Club; We want groups of 2 people. How many possible combinations?

$$\binom{10}{2} = \frac{10!}{2!(10-2)!} = 45$$

Independent Events

- Two events are independent (does not directly affect the probability of the other from happening) iff: P(AB) = P(A)P(B)
- A set of events $\{A_i : i \in I\}$ is independent if $P(\bigcap_{i \in J} A_i) = \prod_{i \in J} P(A_i)$
- Assume A & B are disjoint events, each with +ve probability. Can they be independent?



Ans: **No** because:
Independent = P(AB) = P(A)P(B)
But here P(AB)=φ=0 and P(A)P(B)>0

•Independence is sometimes assumed & sometimes derived.
Ex: Toss a fair die A={2,4,6} B{1,2,3,4} $A \bigcap B = \{2,4\}$
P(AB)=2/6=P(A)P(B)=1/2*2/3

Conditional Probability

- Assuming $P(B)>0 \ldots P(A|B)=P(AB)/P(B)$
- If A & B are *independent* then $P(A|B)=P(A)$
- Ex: Draw Ace of clubs (A) and then Queen of Diamonds (B) $P(AB)=P(A)P(B|A)=1/52*1/51=1/2652$
- $P(.|B)$ satisfies the axioms, for fixed B
- $P(A|.)$ does not satisfy the axioms of probability, for fixed A
- In general $P(A|B) \neq P(B|A)$

Bayes' Theorem

- The Law of Total Probability $$P(B) = \sum_{i=1}^{k} P(B \mid A_i) P(A_i)$$
  - $A_1 \ldots A_k$ are partitions of Ω $$P(B) = \sum P(C_j) = \sum P(BA_j) = \sum P(B \mid A_j) P(A_j)$$

  Where $C_j = BA_j$ $C_{1 \ldots k}$ are disjoint

  - Let $A_1, \ldots, A_k$ be a partition of Ω such that $P(A_i)>0$ for each i. If $P(B)>0$ then, for each $i=1,\ldots,k$

$$P(A_i / B) = \frac{P(B \mid A_i) P(A_i)}{\sum P(B \mid A_j) P(A_j)}$$

Additional Examples

- Disease Test with + or – outcomes

	D	D^c
+	.009	.099
-	.001	.891

Go to test and get + results. What's the probability of one having disease? Ans: not 90%...actually 8%

$$P(D \mid +) = \frac{P(+ \bigcap D)}{P(+)} = \frac{.009}{.009 + .099} \approx 0.08$$

- 3 Catergories of Email: A_1= "spam", A_2= "low priority", A_3= "high priority"
- $P(A_1)=.7, P(A_2)=.2, P(A_3)=.1$ Let B be event that email contains the word "Free"
- $P(B|A_1)=.9$, $P(B|A_2)=.01$, $P(B|A_3)=.01$
- Q: receive an email with word "free", what's the probability that it is spam?

$$P(A_1 \mid B) = \frac{.9 \times .7}{(.9 \times .7) + (.01 \times .2) + (.01 \times .1)} = .995$$

Chapter 2 Overview

- Random Variable
- Cumulative Distribution Function (CDF)
- Discrete Vs. Continuous probability functions (Probability Density Function PDF)
  - Discrete: Point Mass, Discrete Uniform, Bernoulli, Binomial, Geometric, Poisson Distribution
  - Continuous: Uniform, Normal (Gaussian), Exponential, Gamma, Beta, t and Cauchy, X^2
- Bivariate Distribution
- Marginal Distribution
- Independent Random Variables
- Conditional Distribution
- Multivariate Distributions: Multinomial, Multivariate Normal
- Transformations of Random Variables.

Random Variable

- A random variable is a mapping
  - X: Ω → R
  - That assigns a real number X(ω) to each outcome ω
- Ex: Flip coin twice and let X be number of heads.

ω	P(ω)	X(ω)
TT	¼	0
TH	¼	1
HT	¼	1
HH	¼	2

x	P(X=x)
0	¼
1	½
2	¼

Cumulative Distribution Function

- CDF = $F_X: R \to [0,1]$
  - $F_X(x)=P(X \le x)$
- Ex: Flip a fair coin twice and let X be number of heads

$$F_x(x) = \begin{cases} 0 & x<0 \\ 1/4 & 0 \le x<1 \\ 3/4 & 1 \le x<2 \\ 1 & x \ge 2 \end{cases}$$

- Theorem: let X and Y have CDF F and G. If F(x) = G(x) for all x, then $P(X \in A) = P(Y \in A)$
- Theorem: CDF needs to satisfy:
  - F is non-decreasing: $x_1<x_2$, also $F(x_1) \le F(x_2)$
  - F is normalized:

    $$\lim_{x \to -\infty} F(x) = 0$$

    $$\lim_{x \to +\infty} F(x) = 1$$

  - F is right-continuous:

    $F(x) = F(x^+)$ for all,

    where $F(x^+) = \lim_{\substack{y \to x \\ y > x}} F(y)$

Discrete Vs. Continuous Distributions

- Def: X is **discrete** if it take **countably** many values. We define the **probability function** or **probability mass function** for X by $f_X(x)=P(X=x)$
- Def: X is **continuous** if there exists a function f_X such that $f_X(x)\geq 0$ for all x, $\int_{-\infty}^{+\infty} f(x)dx = 1$ and for every a≤b

$$P(a < X < b) = \int_a^b f(x)dx$$

- The function f_X is called the probability density function (PDF).

$$F_x(x) = \int_{-\infty}^{x} fx(t)dt$$

and $f_X(x) = F'_X(x)$ at all points x at which F_X is differentiable

Note: mention Lemmas (pg 25)

Some Important Discrete Functions

- **Point Mass Distribution:** if P(X=a)=1 otherwise 0
- **Discrete Uniform**
  - f(x)= $\begin{cases} \frac{1}{k} & \text{for x=1,...,k} \\ 0 & \text{otherwise} \end{cases}$
- **Bernoulli** (Binary Coin Flip)
  - $f(x)=P^x(1-p)^{1-x}$ for x ε {0,1}
- **Binomial** (flip coins n time and let X be number of heads)

f(x)=P(X=x)={ $\binom{n}{x} P^\wedge x(1-p)^\wedge(n-x)$ for x=0,…,n

{0 otherwise

- **Geometric** p ε (0,1) (Think X as the No. of flips needed to get the 1st head)
  - $P(X=k)=p(1-P)^{k-1}$, k=1,2,3,…
- **Poisson** (Counting rare events like radioactive decay or traffic accidents)

$$f(x) = e^\wedge(-\lambda)\frac{\lambda^\wedge x}{x!} \quad x \geq 0$$

Some Important Continuous Distributions

- **Uniform Distribution**
  - f(x)={1/(b-a) for x ∈ [a,b]
    {0 otherwise
- **Normal (Gaussian)**

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\{-\frac{1}{2\sigma^2}(x-\mu)^2\}$$

Standard Normal if μ=0 and σ=1

if X ~ N(μ, σ^2), then Z=(X- μ)/ σ N(0,1)

- **Exponential** (Model lifetimes of electronic components/ waiting times between rare events)

$$f(x) = \frac{1}{\beta} e^{(-x/\beta)}, x > 0$$

- t distribution is like Normal except with thinker tails
- Cauchy is special case of t distribution where v=1
- X^2 distribution deals with various degrees of freedom

Bivariate Distribution

- Discrete random variables X & Y, define **joint mass function** by f(x,y)=P(X=x and Y=y)
- F(1,1)=P(X=1,Y=1)=4/9
- In continuous case, we use f(x,y) a PDF for the random variables (X,Y)
  - (i) f(x,y) ≥ 0 for all (x,y)
  - (ii) $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy = 1 \ \ and,$
  - for any set A is RXR,P((X,Y) A) = $\iint f(x,y)dxdy$

	Y=0	Y=1	
X=0	1/9	2/9	1/3
X=1	2/9	4/9	2/3
	1/3	2/3	1

Marginal Distributions

- Def: if (X,Y) have joint distribution with mass function $f_{X,Y}$, then the **marginal mass function** for X is defined by

$$f_X(x) = P(X = x) = \sum_y P(X = x, Y = y) = \sum_y f(x, y)$$

- And the marginal Mass Function for Y is defined by

$$f_Y(x) = P(X = x) = \sum_x P(X = x, Y = y) = \sum_x f(x, y)$$

- $f_X(0)=3/10$ and $f_X(1)=7/10$
- For continuous random variables, the marginal densities are

$$f_X(x) = \int f(x, y)dy$$

$$f_Y(y) = \int f(x, y)dx$$

	Y=0	Y=1	
X=0	1/10	2/10	3/10
X=1	3/10	4/10	7/10
	4/10	6/10	

Independent Random Variables

- Def: Two random variables X & Y are **independent** iff, for every A & B,

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$$

	Y=0	Y=1	
X=0	¼	¼	½
X=1	¼	¼	½
	½	½	1

- $f_X(0)f_Y(0)=f(0,0)$
- $f_X(0)f_Y(1)=f(0,1)$
- $f_X(1)f_Y(0)=f(1,0)$
- $f_X(1)f_Y(1)=f(1,1)$

	Y=0	Y=1	
X=0	½	0	½
X=1	0	½	½
	½	½	1

- These are not independent because
- $f_X(0)f_Y(1)=(1/2)(1/2)=1/4$ yet $f(0,1)=0$

Conditional Distributions

- Def:
  - f(x|y)=P(X=x|Y=y)=P(X=x,Y=y)/P(Y=y)=f(x,y)/f(y)
  - Assuming that f(y)>0
- Def: for continuous random variables,
  - f(x|y)= f(x,y)/f(y)
  - Assuming that f(y)>0. Then,

$$P(X \in A / Y = y) = \int_A f(x \mid y) dx$$

Multivariate Distribution and IID Samples

- If $X_1,\ldots,X_n$ are independent and each has the same marginal distribution with CDF F, we say that $X_1,\ldots,X_n$ are independent and identically distributed and we write
  - $X_1,\ldots X_n \sim F$
  - Random sample size n from F

2 Important Multivariable Distributions

- **Multinomial**: multivariate version of a Binomial
  - X ~ Multinomial(n,p)

$$f(x) = \binom{n}{x1 \ldots xk} P1^{\wedge}(x1) \ldots Pk^{\wedge}(xk)$$

Where

$$\binom{n}{x1 \ldots xk} = \frac{n!}{x1! \ldots xk!}$$

- **Multivariate Normal:** μ is a vector and σ is a matrix (pg40)

Transformations of Random Variables

- Three steps for Transformations
  - 1. For each y, find the set $A_y=\{x:r(x)\leq y\}$
  - 2.Find CDF
    - $F(y)=P(Y \leq y)=P(r(X) \leq y)$

      $=P(\{x;r(x) \leq y\})$

      $= \int_{Ay} fx(x)dx$
  - 3. The PDF is $f(y)=F'(y)$