STAB52

Summary Of

Probability and Statistics

The Science of Uncertainty

Second Edition

By: **Michael J. Evans and Jerey S. Rosenthal**

University of Toronto

Instructor:	**Dr. Daniel Roy**
Email:	droy@utsc.utoronto.ca
Office:	IC462
Office Hours:	TU 11:00 - 12:00

1 Probability Basics

1.1 Probability Models

Sample space, often written S. This is any set that lists all possible outcomes of some unknown experiment. Collections of events are subsets of S, to which probabilities can be assigned.
Finally, a probability model requires a probability measure, usually written P. This must assign to each event A, a probability $P(A)$ with the following properties:

1. $P(A)$ is always a non-negative real number, between 0 and 1 inclusive.
2. $P(\varnothing) = 0$
3. $P(S) = 1$
4. P is countably additive, where for disjoint events $A_1, A_2, A_3, \dots$
   we have $P(A_1 \cup A_2 \cup \dots) = P(A_1) + P(A_2) + \dots$

1.2 Venn Diagrams and Subsets

The complement of a set A, denoted set $A^c = \{s | s \notin A\}$
The intersection of two sets A, B, denoted $A \cap B = \{s | s \in A \wedge s \in B\}$
The union of two sets A, B, denoted $A \cup B = \{s | s \in A \wedge s \in B\}$

We also have properties $(A \cup B)^c = A^c \cap B^c$ and $(A \cap B)^c = A^c \cup B^c$

1.3 Properties of Probability Models

For any event A, A and A^c are always disjoint.
Furthermore, their union is always the entire sample space: $A \cup A^c = S$
And since we have $P(S) = 1$. $P(A^c) = 1 - P(A)$

Suppose that $A_1, A_2, \dots$ are disjoint events that form a partition of the sample space i.e., $A_1 \cup A_2 \cup + \dots = S$.
For any event B, $P(B) = P(A_1 \cap B) + P(A_2 \cap B) + \dots$

Principle of inclusion-exclusion, Let A, B be two events. Then
$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

1.4 Uniform Probability on Finite Spaces

If a sample space S is finite, then one possible probability measure on S is the uniform probability measure, which assigns probability $\dfrac{1}{|S|}$ to each outcome. By additivity, we see that for any event $A, P(A) = \dfrac{|A|}{|S|}$

1. Multiplication Principle

   With m in A and n elements in B, there are $m \times n$ total possible ordered pairs of elements from both sets, $C = \{(a_i, b_j) | a_i \in A, b_j \in B\}, |C| = m \times n$

2. Permutation Principle

   Ordered arrangement of k objects, chosen without replacement from n possible objects.

   The number of these ordered arrangements is $P_k^n = \dfrac{n!}{(n-k)!}$

3. Combination Principle

   Unordered arrangement of k objects, chosen without replacement from n possible object.

   The number of these unordered arrangement is $C_k^n = \binom{n}{k} = \frac{P_k^n}{k!} = \frac{n!}{k!(n-k)!}$

1.5 Conditional Probability and Independence

Given two events A, B with $P(B) > 0$, the conditional probability of A given B written $P(A|B)$ denotes the fraction of time that A occurs once we know that B has occured.

$P(A|B) = \frac{P(A \cap B)}{P(B)}$, $P(A \cap B) = P(A)P(B|A)$

Then the law of total probability can be rewritten: Let $A_1, A_2, \dots$ be events that form a partition of the sample space S, each of positive probability.
Then for any event B, $P(B) = P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + \dots$

Let A, B be two events, each of positive probability. Then $P(A|B) = \frac{P(A)P(B|A)}{P(B)}$

Two events A, B are independent if $P(A \cap B) = P(A)P(B)$
Three events A, B, C are independent if **all** of the following equations hold:

1. $P(A \cap B) = P(A)P(B)$
2. $P(A \cap C) = P(A)P(C)$
3. $P(B \cap C) = P(B)P(C)$
4. $P(A \cap B \cap C) = P(A)P(B)P(C)$

If only 1 - 3 hold, then the set is called **pairwise independent**.

2 Random Variables and Distributions

2.1 Random Variables

A random variable is a function from the sample space S to $\mathbb{R}$.

Constant Random Variables
let c be any constant and also also a function, by saying $c(s) = c, \forall s \in S$. Thus, 5 is a random variable, as is 3 or -21.6.

Indicator Functions
If A is any event, then we can define the indicator function I_A to be the random variable such that:

$$I_A(s) = \begin{cases} 1 & s \in A \\ 0 & s \notin A \end{cases}$$

which is equal to 1 on A and is equal to 0 on A^C.

2.2 Distribution of Random Variables

Since random variables are defined to be functions of the outcome s, and because the outcome s is assumed to be random, it follows that the value of a random variable will itself be random.

However, if X is a random variable, then the probability that X will equal to some particular value x is precisely when the outcome of s is chosen such that $X(s) = x$.

If X is a random variable, then the distribution of X is the collection of probabilities $P(X \in B)$ for all subsets B of the real numbers.

2.3 Discrete Distributions

For many random variables X, if we have $P(X = x) > 0$ for certain x values.This means there is positive probability that the variable will be equal to certain particular values.

If $\sum\limits_{x \in \mathbb{R}} P(X = x) = 1$, which says all of the probability assigned with the random variable X sums to 1, this random variable X is discrete.

We can formalize this as: A random variable X is discrete if there is a finite or countable sequence $x_1, x_2, \ldots$ of distinct real numbers, and a corresponding sequence $p_1, p_2, \ldots$ of non-negative real numbers, such that $P(X = x_i) = p_i$ for all i, and $\sum_i p_i = 1$.

For a discrete random variable X, its probability function is the function $P_X : \mathbb{R} \to [0, 1]$ defined by $P_X(y) = P(X = y)$

Distributions

Bernoulli

Consider flipping a coin that has probability θ of coming up heads, and probability of $1 - \theta$ of coming up tails, where $\theta \in [0, 1]$.

Let $X = 1$ of the coin is heads, and $X = 0$ otherwise. then $P_X(1) = P(X = 1) = \theta$ and $P_X(0) = P(X = 0) = 1 - \theta$. The random variable X is said to have the Bernoulli(θ) distribution; we write this as $X \sim$ Bernoulli(θ).

Binomial

Consider flipping n coins, each of which has independent probability of θ of coming up heads, and probability $1 - \theta$ of coming up tails, where $\theta \in [0, 1]$.
Let X be the total number of heads showing, then for each $y = 1, 2, 3, \ldots, n$,
$P_X(y) = P(X = y) = \binom{n}{y}\theta^y(1-\theta)^{n-y} = \dfrac{n!}{(n-y)!y!}\theta^y(1-\theta)^{y-x}$
The random variable X is said to have the Binomial(n, θ) distribution; we write this as $X \sim$ Binomial(n, θ).
The Bernoulli(θ) distribution corresponds to the special case of the Binomial(n, θ) distribution where $n = 1$.

Geometric

Consider repeatedly flipping a coin that has probability θ of coming up heads and probability $1-\theta$ of coming up tails, where again $0 < \theta < 1$. Let X be the number of tails that appear before the first head.

Then for $k \geq 0, X = k$ if and only if the coin shows exactly k tails followed by a head. The probability of this is equal to $(1 - \theta)^k\theta$

Negative-Binomial Distribution

Consider repeatedly flipping a coin that has probability θ of coming up heads and probability $1-\theta$ of coming up tails. Let r be a positive integer, and let Y be the number of tails that appear before the r-th head.

Then for $k \geq 0, Y = k$ if and only if the coin shows exactly $r-1$ heads and k tails on the first $r+k-1$ flips, then shows a head on the $r+k$-th flip. The probability of this is equal to:

$$P_Y(k) = \binom{r+k-1}{r-1}\theta^{r-1}(1-\theta)^k\theta = \binom{r+k-1}{r-1}\theta^r(1-\theta)^k$$

The random variable Y is said to have the Negative-Binomial(r,θ) distribution; we write this as $Y \sim$ Negative-Binomial(r,θ). Of course, the special case $r=1$ is the Geometric(θ) distribution.

The Poisson Distribution

We say that if a random variable Y has the poisson distribution, and write $Y \sim$ Poisson(λ), if

$$P_Y(x) = P(Y = x) = \frac{\lambda^y}{y!}e^{-\lambda}$$

For $y = 0, 1, 2, 3, \ldots$ We note that since $\sum_{y=0}^{\infty} \lambda^y/y! = e^y$ (From Calculus), then $\sum_{y=0}^{\infty} P(Y = y) = 1$

We motivate the Poisson distribution as follows. Suppose $X \sim$ Binomial(n,θ), then for $0 \leq x \leq n$ $P(X = x) = \binom{n}{x}\theta^x(1-\theta)^{n-x}$,

The Hyper-geometric Distribution

Suppose an urn contains N total balls, M white balls and $N - M$ black balls.

2.4 Continuous Distributions

A random variable X is continuous if $P(X = x) = 0, \forall x \in \mathbb{R}$

We say that a random variable X is absolutely continuous if there is a density function f, such that $P(X \in [a,b]) = \int_a^b f(x)dx$ whenever $a \leq b$

Important Absolutely Continuous Distributions

The Uniform[0,1] Distribution