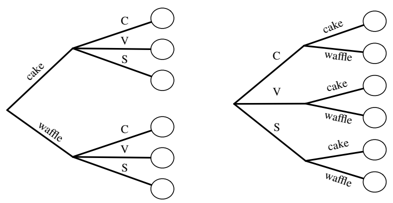
Probability Cheatsheet v2.0

Compiled by William Chen (http://wzchen.com) and Joe Blitzstein, with contributions from Sebastian Chiu, Yuan Jiang, Yuqi Hou, and Jessy Hwang. Material based on Joe Blitzstein's (@stat110) lectures (http://stat110.net) and Blitzstein/Hwang's Introduction to Probability textbook (http://bit.ly/introprobability). Please share comments, suggestions, and errors at http://github.com/wzchen/probability_cheatsheet.

Last Updated September 4, 2015

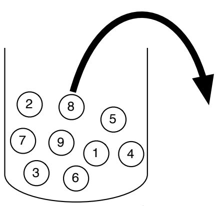
Counting

Multiplication Rule



Let's say we have a compound experiment (an experiment with multiple components). If the 1st component has n_1 possible outcomes, the 2nd component has n_2 possible outcomes, $\dots$, and the rth component has n_r possible outcomes, then overall there are $n_1 n_2 \dots n_r$ possibilities for the whole experiment.

Sampling Table



The sampling table gives the number of possible samples of size k out of a population of size n, under various assumptions about how the sample is collected.

	Order Matters	**Not Matter**
With Replacement	n^k	$\binom{n+k-1}{k}$
Without Replacement	$\frac{n!}{(n-k)!}$	$\binom{n}{k}$

Naive Definition of Probability

If all outcomes are equally likely, the probability of an event A happening is:

$$P_{\text{naive}}(A) = \frac{\text{number of outcomes favorable to } A}{\text{number of outcomes}}$$

Thinking Conditionally

Independence

Independent Events A and B are independent if knowing whether A occurred gives no information about whether B occurred. More formally, A and B (which have nonzero probability) are independent if and only if one of the following equivalent statements holds:

$$P(A \cap B) = P(A)P(B)$$
$$P(A|B) = P(A)$$
$$P(B|A) = P(B)$$

Conditional Independence A and B are conditionally independent given C if $P(A \cap B|C) = P(A|C)P(B|C)$. Conditional independence does not imply independence, and independence does not imply conditional independence.

Unions, Intersections, and Complements

De Morgan's Laws A useful identity that can make calculating probabilities of unions easier by relating them to intersections, and vice versa. Analogous results hold with more than two sets.

$$(A \cup B)^c = A^c \cap B^c$$
$$(A \cap B)^c = A^c \cup B^c$$

Joint, Marginal, and Conditional

Joint Probability $P(A \cap B)$ or $P(A, B)$ – Probability of A and B.

Marginal (Unconditional) Probability $P(A)$ – Probability of A.

Conditional Probability $P(A|B) = P(A,B)/P(B)$ – Probability of A, given that B occurred.

Conditional Probability *is* Probability $P(A|B)$ is a probability function for any fixed B. Any theorem that holds for probability also holds for conditional probability.

Probability of an Intersection or Union

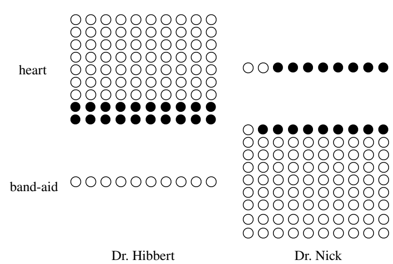
Intersections via Conditioning

$$P(A, B) = P(A)P(B|A)$$
$$P(A, B, C) = P(A)P(B|A)P(C|A, B)$$

Unions via Inclusion-Exclusion

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$
$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C).$$

Simpson's Paradox



It is possible to have

$$P(A \mid B, C) < P(A \mid B^c, C) \text{ and } P(A \mid B, C^c) < P(A \mid B^c, C^c)$$
$$\text{yet also } P(A \mid B) > P(A \mid B^c).$$

Law of Total Probability (LOTP)

Let $B_1, B_2, B_3, \dots B_n$ be a *partition* of the sample space (i.e., they are disjoint and their union is the entire sample space).

$$P(A) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + \dots + P(A|B_n)P(B_n)$$
$$P(A) = P(A \cap B_1) + P(A \cap B_2) + \dots + P(A \cap B_n)$$

For **LOTP with extra conditioning**, just add in another event C!

$$P(A|C) = P(A|B_1, C)P(B_1|C) + \dots + P(A|B_n, C)P(B_n|C)$$
$$P(A|C) = P(A \cap B_1|C) + P(A \cap B_2|C) + \dots + P(A \cap B_n|C)$$

Special case of LOTP with B and B^c as partition:

$$P(A) = P(A|B)P(B) + P(A|B^c)P(B^c)$$
$$P(A) = P(A \cap B) + P(A \cap B^c)$$

Bayes' Rule

Bayes' Rule, and with extra conditioning (just add in C!)

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

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

We can also write

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

Odds Form of Bayes' Rule

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

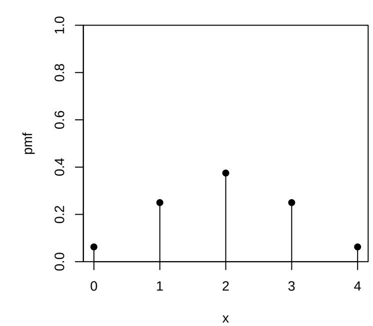
The *posterior odds* of A are the *likelihood ratio* times the *prior odds*.

Random Variables and their Distributions

PMF, CDF, and Independence

Probability Mass Function (PMF) Gives the probability that a *discrete* random variable takes on the value x.

$$p_X(x) = P(X = x)$$

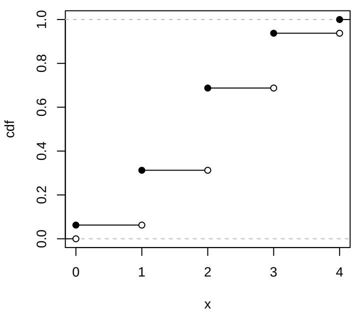


The PMF satisfies

$$p_X(x) \geq 0 \text{ and } \sum_x p_X(x) = 1$$

Cumulative Distribution Function (CDF) Gives the probability that a random variable is less than or equal to x.

$$F_X(x) = P(X \leq x)$$



The CDF is an increasing, right-continuous function with

$$F_X(x) \to 0 \text{ as } x \to -\infty \text{ and } F_X(x) \to 1 \text{ as } x \to \infty$$

Independence Intuitively, two random variables are independent if knowing the value of one gives no information about the other. Discrete r.v.s X and Y are independent if for *all* values of x and y

$$P(X = x, Y = y) = P(X = x)P(Y = y)$$

Expected Value and Indicators

Expected Value and Linearity

Expected Value (a.k.a. *mean*, *expectation*, or *average*) is a weighted average of the possible outcomes of our random variable. Mathematically, if $x_1, x_2, x_3, \ldots$ are all of the distinct possible values that X can take, the expected value of X is

$$E(X) = \sum_i x_i P(X = x_i)$$

X	Y	$X+Y$
3	4	7
2	2	4
6	8	14
10	23	33
1	–3	–2
1	0	1
5	9	14
4	1	5
...	...	...

$$\frac{1}{n}\sum_{i=1}^{n} x_i \quad + \quad \frac{1}{n}\sum_{i=1}^{n} y_i \quad = \quad \frac{1}{n}\sum_{i=1}^{n} (x_i + y_i)$$

$$E(X) \quad + \quad E(Y) \quad = \quad E(X+Y)$$

Linearity For any r.v.s X and Y, and constants a, b, c,

$$E(aX + bY + c) = aE(X) + bE(Y) + c$$

Same distribution implies same mean If X and Y have the same distribution, then $E(X) = E(Y)$ and, more generally,

$$E(g(X)) = E(g(Y))$$

Conditional Expected Value is defined like expectation, only conditioned on any event A.

$$E(X|A) = \sum_x xP(X = x|A)$$

Indicator Random Variables

Indicator Random Variable is a random variable that takes on the value 1 or 0. It is always an indicator of some event: if the event occurs, the indicator is 1; otherwise it is 0. They are useful for many problems about counting how many events of some kind occur. Write

$$I_A = \begin{cases} 1 & \text{if } A \text{ occurs,} \\ 0 & \text{if } A \text{ does not occur.} \end{cases}$$

Note that $I_A^2 = I_A, I_A I_B = I_{A\cap B}$, and $I_{A\cup B} = I_A + I_B - I_A I_B$.

Distribution $I_A \sim \text{Bern}(p)$ where $p = P(A)$.

Fundamental Bridge The expectation of the indicator for event A is the probability of event A: $E(I_A) = P(A)$.

Variance and Standard Deviation

$$\text{Var}(X) = E\left(X - E(X)\right)^2 = E(X^2) - (E(X))^2$$

$$\text{SD}(X) = \sqrt{\text{Var}(X)}$$

Continuous RVs, LOTUS, UoU

Continuous Random Variables (CRVs)

What's the probability that a CRV is in an interval? Take the difference in CDF values (or use the PDF as described later).

$$P(a \leq X \leq b) = P(X \leq b) - P(X \leq a) = F_X(b) - F_X(a)$$

For $X \sim \mathcal{N}(\mu, \sigma^2)$, this becomes

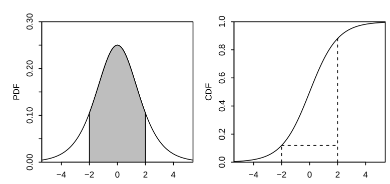
$$P(a \leq X \leq b) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

What is the Probability Density Function (PDF)? The PDF f is the derivative of the CDF F.

$$F'(x) = f(x)$$

A PDF is nonnegative and integrates to 1. By the fundamental theorem of calculus, to get from PDF back to CDF we can integrate:

$$F(x) = \int_{-\infty}^{x} f(t)dt$$



To find the probability that a CRV takes on a value in an interval, integrate the PDF over that interval.

$$F(b) - F(a) = \int_a^b f(x)dx$$

How do I find the expected value of a CRV? Analogous to the discrete case, where you sum x times the PMF, for CRVs you integrate x times the PDF.

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx$$

LOTUS

Expected value of a function of an r.v. The expected value of X is defined this way:

$$E(X) = \sum_x xP(X = x) \text{ (for discrete } X\text{)}$$

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx \text{ (for continuous } X\text{)}$$

The **Law of the Unconscious Statistician (LOTUS)** states that you can find the expected value of a *function of a random variable*, $g(X)$, in a similar way, by replacing the x in front of the PMF/PDF by $g(x)$ but still working with the PMF/PDF of X:

$$E(g(X)) = \sum_x g(x)P(X = x) \text{ (for discrete } X\text{)}$$

$$E(g(X)) = \int_{-\infty}^{\infty} g(x)f(x)dx \text{ (for continuous } X\text{)}$$

What's a function of a random variable? A function of a random variable is also a random variable. For example, if X is the number of bikes you see in an hour, then $g(X) = 2X$ is the number of bike wheels you see in that hour and $h(X) = \binom{X}{2} = \frac{X(X-1)}{2}$ is the number of *pairs* of bikes such that you see both of those bikes in that hour.

What's the point? You don't need to know the PMF/PDF of $g(X)$ to find its expected value. All you need is the PMF/PDF of X.

Universality of Uniform (UoU)

When you plug any CRV into its own CDF, you get a Uniform(0,1) random variable. When you plug a Uniform(0,1) r.v. into an inverse CDF, you get an r.v. with that CDF. For example, let's say that a random variable X has CDF

$$F(x) = 1 - e^{-x}, \text{ for } x > 0$$

By UoU, if we plug X into this function then we get a uniformly distributed random variable.

$$F(X) = 1 - e^{-X} \sim \text{Unif}(0, 1)$$

Similarly, if $U \sim \text{Unif}(0, 1)$ then $F^{-1}(U)$ has CDF F. The key point is that for any continuous random variable X, we can transform it into a Uniform random variable and back by using its CDF.

Moments and MGFs

Moments

Moments describe the shape of a distribution. Let X have mean μ and standard deviation σ, and $Z = (X - \mu)/\sigma$ be the *standardized* version of X. The kth moment of X is $\mu_k = E(X^k)$ and the kth standardized moment of X is $m_k = E(Z^k)$. The mean, variance, skewness, and kurtosis are important summaries of the shape of a distribution.

Mean $E(X) = \mu_1$

Variance $\text{Var}(X) = \mu_2 - \mu_1^2$

Skewness $\text{Skew}(X) = m_3$

Kurtosis $\text{Kurt}(X) = m_4 - 3$

Moment Generating Functions

MGF For any random variable X, the function

$$M_X(t) = E(e^{tX})$$

is the **moment generating function (MGF)** of X, if it exists for all t in some open interval containing 0. The variable t could just as well have been called u or v. It's a bookkeeping device that lets us work with the *function* M_X rather than the *sequence* of moments.

Why is it called the Moment Generating Function? Because the kth derivative of the moment generating function, evaluated at 0, is the kth moment of X.

$$\mu_k = E(X^k) = M_X^{(k)}(0)$$

This is true by Taylor expansion of e^{tX} since

$$M_X(t) = E(e^{tX}) = \sum_{k=0}^{\infty} \frac{E(X^k)t^k}{k!} = \sum_{k=0}^{\infty} \frac{\mu_k t^k}{k!}$$

MGF of linear functions If we have $Y = aX + b$, then

$$M_Y(t) = E(e^{t(aX+b)}) = e^{bt}E(e^{(at)X}) = e^{bt}M_X(at)$$

Uniqueness *If it exists, the MGF uniquely determines the distribution.* This means that for any two random variables X and Y, they are distributed the same (their PMFs/PDFs are equal) if and only if their MGFs are equal.

Summing Independent RVs by Multiplying MGFs. If X and Y are independent, then

$$M_{X+Y}(t) = E(e^{t(X+Y)}) = E(e^{tX})E(e^{tY}) = M_X(t) \cdot M_Y(t)$$

The MGF of the sum of two random variables is the product of the MGFs of those two random variables.

Joint PDFs and CDFs

Joint Distributions

The **joint CDF** of X and Y is

$$F(x, y) = P(X \le x, Y \le y)$$

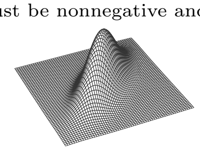
In the discrete case, X and Y have a **joint PMF**

$$p_{X,Y}(x, y) = P(X = x, Y = y).$$

In the continuous case, they have a **joint PDF**

$$f_{X,Y}(x, y) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x, y).$$

The joint PMF/PDF must be nonnegative and sum/integrate to 1.



Conditional Distributions

Conditioning and Bayes' rule for discrete r.v.s

$$P(Y = y|X = x) = \frac{P(X = x, Y = y)}{P(X = x)} = \frac{P(X = x|Y = y)P(Y = y)}{P(X = x)}$$

Conditioning and Bayes' rule for continuous r.v.s

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x, y)}{f_X(x)} = \frac{f_{X|Y}(x|y)f_Y(y)}{f_X(x)}$$

Hybrid Bayes' rule

$$f_X(x|A) = \frac{P(A|X = x)f_X(x)}{P(A)}$$

Marginal Distributions

To find the distribution of one (or more) random variables from a joint PMF/PDF, sum/integrate over the unwanted random variables.

Marginal PMF from joint PMF

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

Marginal PDF from joint PDF

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)dy$$

Independence of Random Variables

Random variables X and Y are independent if and only if any of the following conditions holds:

- Joint CDF is the product of the marginal CDFs
- Joint PMF/PDF is the product of the marginal PMFs/PDFs
- Conditional distribution of Y given X is the marginal distribution of Y

Write $X \perp\!\!\!\perp Y$ to denote that X and Y are independent.

Multivariate LOTUS

LOTUS in more than one dimension is analogous to the 1D LOTUS. For discrete random variables:

$$E(g(X, Y)) = \sum_x \sum_y g(x, y)P(X = x, Y = y)$$

For continuous random variables:

$$E(g(X, Y)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y)f_{X,Y}(x, y)dxdy$$

Covariance and Transformations

Covariance and Correlation

Covariance is the analog of variance for two random variables.

$$\text{Cov}(X, Y) = E\left((X - E(X))(Y - E(Y))\right) = E(XY) - E(X)E(Y)$$

Note that

$$\text{Cov}(X, X) = E(X^2) - (E(X))^2 = \text{Var}(X)$$

Correlation is a standardized version of covariance that is always between -1 and 1.

$$\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}$$

Covariance and Independence If two random variables are independent, then they are uncorrelated. The converse is not necessarily true (e.g., consider $X \sim \mathcal{N}(0, 1)$ and $Y = X^2$).

$$X \perp\!\!\!\perp Y \longrightarrow \text{Cov}(X, Y) = 0 \longrightarrow E(XY) = E(X)E(Y)$$

Covariance and Variance The variance of a sum can be found by

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y)$$

$$\text{Var}(X_1 + X_2 + \cdots + X_n) = \sum_{i=1}^{n} \text{Var}(X_i) + 2\sum_{i<j} \text{Cov}(X_i, X_j)$$

If X and Y are independent then they have covariance 0, so

$$X \perp\!\!\!\perp Y \Longrightarrow \text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$$

If $X_1, X_2, \dots, X_n$ are identically distributed and have the same covariance relationships (often by **symmetry**), then

$$\text{Var}(X_1 + X_2 + \cdots + X_n) = n\text{Var}(X_1) + 2\binom{n}{2}\text{Cov}(X_1, X_2)$$

Covariance Properties For random variables W, X, Y, Z and constants a, b:

$$\begin{aligned}
\text{Cov}(X, Y) &= \text{Cov}(Y, X) \\
\text{Cov}(X + a, Y + b) &= \text{Cov}(X, Y) \\
\text{Cov}(aX, bY) &= ab\text{Cov}(X, Y) \\
\text{Cov}(W + X, Y + Z) &= \text{Cov}(W, Y) + \text{Cov}(W, Z) + \text{Cov}(X, Y) \\
&\quad + \text{Cov}(X, Z)
\end{aligned}$$

Correlation is location-invariant and scale-invariant For any constants a, b, c, d with a and c nonzero,

$$\text{Corr}(aX + b, cY + d) = \text{Corr}(X, Y)$$

Transformations

One Variable Transformations Let's say that we have a random variable X with PDF $f_X(x)$, but we are also interested in some function of X. We call this function $Y = g(X)$. Also let $y = g(x)$. If g is differentiable and strictly increasing (or strictly decreasing), then the PDF of Y is

$$f_Y(y) = f_X(x)\left|\frac{dx}{dy}\right| = f_X(g^{-1}(y))\left|\frac{d}{dy}g^{-1}(y)\right|$$

The derivative of the inverse transformation is called the **Jacobian**.

Two Variable Transformations Similarly, let's say we know the joint PDF of U and V but are also interested in the random vector (X, Y) defined by $(X, Y) = g(U, V)$. Let

$$\frac{\partial(u, v)}{\partial(x, y)} = \begin{pmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{pmatrix}$$

be the **Jacobian matrix.** If the entries in this matrix exist and are continuous, and the determinant of the matrix is never 0, then

$$f_{X,Y}(x, y) = f_{U,V}(u, v)\left\|\frac{\partial(u, v)}{\partial(x, y)}\right\|$$

The inner bars tells us to take the matrix's determinant, and the outer bars tell us to take the absolute value. In a 2×2 matrix,

$$\left\| \begin{matrix} a & b \\ c & d \end{matrix} \right\| = |ad - bc|$$

Convolutions

Convolution Integral If you want to find the PDF of the sum of two independent CRVs X and Y, you can do the following integral:

$$f_{X+Y}(t) = \int_{-\infty}^{\infty} f_X(x)f_Y(t - x)dx$$

Example Let $X, Y \sim \mathcal{N}(0, 1)$ be i.i.d. Then for each fixed t,

$$f_{X+Y}(t) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}e^{-x^2/2}\frac{1}{\sqrt{2\pi}}e^{-(t-x)^2/2}dx$$

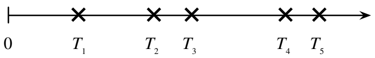
By completing the square and using the fact that a Normal PDF integrates to 1, this works out to $f_{X+Y}(t)$ being the $\mathcal{N}(0, 2)$ PDF.

Poisson Process

Definition We have a **Poisson process** of rate λ arrivals per unit time if the following conditions hold:

1. The number of arrivals in a time interval of length t is $\text{Pois}(\lambda t)$.
2. Numbers of arrivals in disjoint time intervals are independent.

For example, the numbers of arrivals in the time intervals $[0, 5]$, $(5, 12)$, and $[13, 23)$ are independent with $\text{Pois}(5\lambda), \text{Pois}(7\lambda), \text{Pois}(10\lambda)$ distributions, respectively.



Count-Time Duality Consider a Poisson process of emails arriving in an inbox at rate λ emails per hour. Let T_n be the time of arrival of the nth email (relative to some starting time 0) and N_t be the number of emails that arrive in $[0, t]$. Let's find the distribution of T_1. The event $T_1 > t$, the event that you have to wait more than t hours to get the first email, is the same as the event $N_t = 0$, which is the event that there are no emails in the first t hours. So

$$P(T_1 > t) = P(N_t = 0) = e^{-\lambda t} \longrightarrow P(T_1 \le t) = 1 - e^{-\lambda t}$$

Thus we have $T_1 \sim \text{Expo}(\lambda)$. By the memoryless property and similar reasoning, the interarrival times between emails are i.i.d. $\text{Expo}(\lambda)$, i.e., the differences $T_n - T_{n-1}$ are i.i.d. $\text{Expo}(\lambda)$.

Order Statistics

Definition Let's say you have n i.i.d. r.v.s $X_1, X_2, \dots, X_n$. If you arrange them from smallest to largest, the ith element in that list is the ith order statistic, denoted $X_{(i)}$. So $X_{(1)}$ is the smallest in the list and $X_{(n)}$ is the largest in the list.

Note that the order statistics are *dependent*, e.g., learning $X_{(4)} = 42$ gives us the information that $X_{(1)}, X_{(2)}, X_{(3)}$ are ≤ 42 and $X_{(5)}, X_{(6)}, \dots, X_{(n)}$ are ≥ 42.

Distribution Taking n i.i.d. random variables $X_1, X_2, \dots, X_n$ with CDF $F(x)$ and PDF $f(x)$, the CDF and PDF of $X_{(i)}$ are:

$$F_{X_{(i)}}(x) = P(X_{(i)} \le x) = \sum_{k=i}^{n} \binom{n}{k} F(x)^k (1 - F(x))^{n-k}$$

$$f_{X_{(i)}}(x) = n \binom{n-1}{i-1} F(x)^{i-1} (1 - F(x))^{n-i} f(x)$$

Uniform Order Statistics The jth order statistic of i.i.d. $U_1, \dots, U_n \sim \text{Unif}(0, 1)$ is $U_{(j)} \sim \text{Beta}(j, n - j + 1)$.

Conditional Expectation

Conditioning on an Event We can find $E(Y|A)$, the expected value of Y given that event A occurred. A very important case is when A is the event $X = x$. Note that $E(Y|A)$ is a *number*. For example:

- The expected value of a fair die roll, given that it is prime, is $\frac{1}{3} \cdot 2 + \frac{1}{3} \cdot 3 + \frac{1}{3} \cdot 5 = \frac{10}{3}$.
- Let Y be the number of successes in 10 independent Bernoulli trials with probability p of success. Let A be the event that the first 3 trials are all successes. Then

  $$E(Y|A) = 3 + 7p$$

  since the number of successes among the last 7 trials is $\text{Bin}(7, p)$.

- Let $T \sim \text{Expo}(1/10)$ be how long you have to wait until the shuttle comes. Given that you have already waited t minutes, the expected additional waiting time is 10 more minutes, by the memoryless property. That is, $E(T|T > t) = t + 10$.

Discrete Y	**Continuous** Y
$E(Y) = \sum_y yP(Y = y)$	$E(Y) = \int_{-\infty}^{\infty} y f_Y(y) dy$
$E(Y|A) = \sum_y yP(Y = y|A)$	$E(Y|A) = \int_{-\infty}^{\infty} y f(y|A) dy$

Conditioning on a Random Variable We can also find $E(Y|X)$, the expected value of Y given the random variable X. This is *a function of the random variable* X. It is *not* a number except in certain special cases such as if $X \perp\!\!\!\perp Y$. To find $E(Y|X)$, find $E(Y|X = x)$ and then plug in X for x. For example:

- If $E(Y|X = x) = x^3 + 5x$, then $E(Y|X) = X^3 + 5X$.
- Let Y be the number of successes in 10 independent Bernoulli trials with probability p of success and X be the number of successes among the first 3 trials. Then $E(Y|X) = X + 7p$.
- Let $X \sim \mathcal{N}(0, 1)$ and $Y = X^2$. Then $E(Y|X = x) = x^2$ since if we know $X = x$ then we know $Y = x^2$. And $E(X|Y = y) = 0$ since if we know $Y = y$ then we know $X = \pm\sqrt{y}$, with equal probabilities (by symmetry). So $E(Y|X) = X^2, E(X|Y) = 0$.

Properties of Conditional Expectation

1. $E(Y|X) = E(Y)$ if $X \perp\!\!\!\perp Y$
2. $E(h(X)W|X) = h(X)E(W|X)$ (**taking out what's known**) In particular, $E(h(X)|X) = h(X)$.
3. $E(E(Y|X)) = E(Y)$ (**Adam's Law**, a.k.a. Law of Total Expectation)

Adam's Law (a.k.a. Law of Total Expectation) can also be written in a way that looks analogous to LOTP. For any events $A_1, A_2, \dots, A_n$ that partition the sample space,

$$E(Y) = E(Y|A_1)P(A_1) + \dots + E(Y|A_n)P(A_n)$$

For the special case where the partition is A, A^c, this says

$$E(Y) = E(Y|A)P(A) + E(Y|A^c)P(A^c)$$

Eve's Law (a.k.a. Law of Total Variance)

$$\text{Var}(Y) = E(\text{Var}(Y|X)) + \text{Var}(E(Y|X))$$

MVN, LLN, CLT

Law of Large Numbers (LLN)

Let $X_1, X_2, X_3 \dots$ be i.i.d. with mean μ. The **sample mean** is

$$\bar{X}_n = \frac{X_1 + X_2 + X_3 + \dots + X_n}{n}$$

The **Law of Large Numbers** states that as $n \to \infty$, $\bar{X}_n \to \mu$ with probability 1. For example, in flips of a coin with probability p of Heads, let X_j be the indicator of the jth flip being Heads. Then LLN says the proportion of Heads converges to p (with probability 1).

Central Limit Theorem (CLT)

Approximation using CLT

We use $\dot{\sim}$ to denote *is approximately distributed*. We can use the **Central Limit Theorem** to approximate the distribution of a random variable $Y = X_1 + X_2 + \dots + X_n$ that is a sum of n i.i.d. random variables X_i. Let $E(Y) = \mu_Y$ and $\text{Var}(Y) = \sigma_Y^2$. The CLT says

$$Y \dot{\sim} \mathcal{N}(\mu_Y, \sigma_Y^2)$$

If the X_i are i.i.d. with mean μ_X and variance σ_X^2, then $\mu_Y = n\mu_X$ and $\sigma_Y^2 = n\sigma_X^2$. For the sample mean $\bar{X}_n$, the CLT says

$$\bar{X}_n = \frac{1}{n}(X_1 + X_2 + \dots + X_n) \dot{\sim} \mathcal{N}(\mu_X, \sigma_X^2/n)$$

Asymptotic Distributions using CLT

We use $\xrightarrow{D}$ to denote *converges in distribution to* as $n \to \infty$. The CLT says that if we standardize the sum $X_1 + \dots + X_n$ then the distribution of the sum converges to $\mathcal{N}(0, 1)$ as $n \to \infty$:

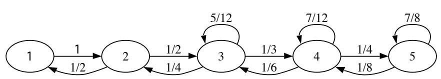
$$\frac{1}{\sigma\sqrt{n}}(X_1 + \dots + X_n - n\mu_X) \xrightarrow{D} \mathcal{N}(0, 1)$$

In other words, the CDF of the left-hand side goes to the standard Normal CDF, Φ. In terms of the sample mean, the CLT says

$$\frac{\sqrt{n}(\bar{X}_n - \mu_X)}{\sigma_X} \xrightarrow{D} \mathcal{N}(0, 1)$$

Markov Chains

Definition



A Markov chain is a random walk in a **state space**, which we will assume is finite, say $\{1, 2, \dots, M\}$. We let X_t denote which element of the state space the walk is visiting at time t. The Markov chain is the sequence of random variables tracking where the walk is at all points in time, $X_0, X_1, X_2, \dots$. By definition, a Markov chain must satisfy the **Markov property**, which says that if you want to predict where the chain will be at a future time, if we know the present state then the entire past history is irrelevant. *Given the present, the past and future are conditionally independent.* In symbols,

$$P(X_{n+1} = j|X_0 = i_0, X_1 = i_1, \dots, X_n = i) = P(X_{n+1} = j|X_n = i)$$

State Properties

A state is either recurrent or transient.

- If you start at a **recurrent state**, then you will always return back to that state at some point in the future. ♪*You can check-out any time you like, but you can never leave.* ♪
- Otherwise you are at a **transient state**. There is some positive probability that once you leave you will never return. ♪*You don't have to go home, but you can't stay here.* ♪

A state is either periodic or aperiodic.

- If you start at a **periodic state** of period k, then the GCD of the possible numbers of steps it would take to return back is $k > 1$.
- Otherwise you are at an **aperiodic state**. The GCD of the possible numbers of steps it would take to return back is 1.

Transition Matrix

Let the state space be $\{1, 2, \dots, M\}$. The transition matrix Q is the $M \times M$ matrix where element q_{ij} is the probability that the chain goes from state i to state j in one step:

$$q_{ij} = P(X_{n+1} = j | X_n = i)$$

To find the probability that the chain goes from state i to state j in exactly m steps, take the (i, j) element of Q^m.

$$q_{ij}^{(m)} = P(X_{n+m} = j | X_n = i)$$

If X_0 is distributed according to the row vector PMF $\vec{p}$, i.e., $p_j = P(X_0 = j)$, then the PMF of X_n is $\vec{p}Q^n$.

Chain Properties

A chain is **irreducible** if you can get from anywhere to anywhere. If a chain (on a finite state space) is irreducible, then all of its states are recurrent. A chain is **periodic** if any of its states are periodic, and is **aperiodic** if none of its states are periodic. In an irreducible chain, all states have the same period.

A chain is **reversible** with respect to $\vec{s}$ if $s_i q_{ij} = s_j q_{ji}$ for all i, j. Examples of reversible chains include any chain with $q_{ij} = q_{ji}$, with $\vec{s} = (\frac{1}{M}, \frac{1}{M}, \dots, \frac{1}{M})$, and random walk on an undirected network.

Stationary Distribution

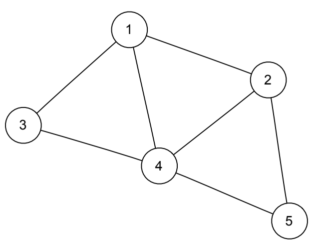
Let us say that the vector $\vec{s} = (s_1, s_2, \dots, s_M)$ be a PMF (written as a row vector). We will call $\vec{s}$ the **stationary distribution** for the chain if $\vec{s}Q = \vec{s}$. As a consequence, if X_t has the stationary distribution, then all future $X_{t+1}, X_{t+2}, \dots$ also have the stationary distribution.

For irreducible, aperiodic chains, the stationary distribution exists, is unique, and s_i is the long-run probability of a chain being at state i. The expected number of steps to return to i starting from i is $1/s_i$.

To find the stationary distribution, you can solve the matrix equation $(Q' - I)\vec{s}\,' = 0$. The stationary distribution is uniform if the columns of Q sum to 1.

Reversibility Condition Implies Stationarity If you have a PMF $\vec{s}$ and a Markov chain with transition matrix Q, then $s_i q_{ij} = s_j q_{ji}$ for all states i, j implies that $\vec{s}$ is stationary.

Random Walk on an Undirected Network



If you have a collection of **nodes**, pairs of which can be connected by undirected **edges**, and a Markov chain is run by going from the current node to a uniformly random node that is connected to it by an edge, then this is a random walk on an undirected network. The stationary distribution of this chain is proportional to the **degree sequence** (this is the sequence of degrees, where the degree of a node is how many edges are attached to it). For example, the stationary distribution of random walk on the network shown above is proportional to $(3, 3, 2, 4, 2)$, so it's $(\frac{3}{14}, \frac{3}{14}, \frac{3}{14}, \frac{4}{14}, \frac{2}{14})$.

Continuous Distributions

Uniform Distribution

Let us say that U is distributed $\text{Unif}(a, b)$. We know the following:

Properties of the Uniform For a Uniform distribution, the probability of a draw from any interval within the support is proportional to the length of the interval. See *Universality of Uniform* and *Order Statistics* for other properties.

Example William throws darts really badly, so his darts are uniform over the whole room because they're equally likely to appear anywhere. William's darts have a Uniform distribution on the surface of the room. The Uniform is the only distribution where the probability of hitting in any specific region is proportional to the length/area/volume of that region, and where the density of occurrence in any one specific spot is constant throughout the whole support.

Normal Distribution

Let us say that X is distributed $\mathcal{N}(\mu, \sigma^2)$. We know the following:

Central Limit Theorem The Normal distribution is ubiquitous because of the Central Limit Theorem, which states that the sample mean of i.i.d. r.v.s will approach a Normal distribution as the sample size grows, regardless of the initial distribution.

Location-Scale Transformation Every time we shift a Normal r.v. (by adding a constant) or rescale a Normal (by multiplying by a constant), we change it to another Normal r.v. For any Normal $X \sim \mathcal{N}(\mu, \sigma^2)$, we can transform it to the standard $\mathcal{N}(0, 1)$ by the following transformation:

$$Z = \frac{X - \mu}{\sigma} \sim \mathcal{N}(0, 1)$$

Standard Normal The Standard Normal, $Z \sim \mathcal{N}(0, 1)$, has mean 0 and variance 1. Its CDF is denoted by Φ.

Exponential Distribution

Let us say that X is distributed $\text{Expo}(\lambda)$. We know the following:

Story You're sitting on an open meadow right before the break of dawn, wishing that airplanes in the night sky were shooting stars, because you could really use a wish right now. You know that shooting stars come on average every 15 minutes, but a shooting star is not "due" to come just because you've waited so long. Your waiting time is memoryless; the additional time until the next shooting star comes does not depend on how long you've waited already.

Example The waiting time until the next shooting star is distributed $\text{Expo}(4)$ hours. Here $\lambda = 4$ is the **rate parameter**, since shooting stars arrive at a rate of 1 per 1/4 hour on average. The expected time until the next shooting star is $1/\lambda = 1/4$ hour.

Expos as a rescaled Expo(1)

$$Y \sim \text{Expo}(\lambda) \rightarrow X = \lambda Y \sim \text{Expo}(1)$$

Memorylessness The Exponential Distribution is the only continuous memoryless distribution. The memoryless property says that for $X \sim \text{Expo}(\lambda)$ and any positive numbers s and t,

$$P(X > s + t | X > s) = P(X > t)$$

Equivalently,

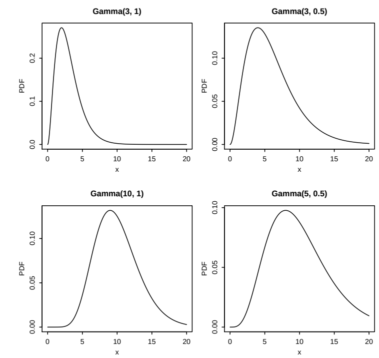
$$X - a | (X > a) \sim \text{Expo}(\lambda)$$

For example, a product with an $\text{Expo}(\lambda)$ lifetime is always "as good as new" (it doesn't experience wear and tear). Given that the product has survived a years, the additional time that it will last is still $\text{Expo}(\lambda)$.

Min of Expos If we have independent $X_i \sim \text{Expo}(\lambda_i)$, then $\min(X_1, \dots, X_k) \sim \text{Expo}(\lambda_1 + \lambda_2 + \dots + \lambda_k)$.

Max of Expos If we have i.i.d. $X_i \sim \text{Expo}(\lambda)$, then $\max(X_1, \dots, X_k)$ has the same distribution as $Y_1 + Y_2 + \dots + Y_k$, where $Y_j \sim \text{Expo}(j\lambda)$ and the Y_j are independent.

Gamma Distribution

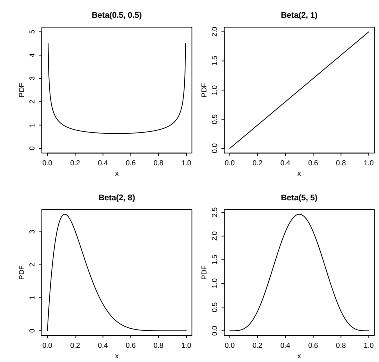


Let us say that X is distributed $\text{Gamma}(a, \lambda)$. We know the following:

Story You sit waiting for shooting stars, where the waiting time for a star is distributed $\text{Expo}(\lambda)$. You want to see n shooting stars before you go home. The total waiting time for the nth shooting star is $\text{Gamma}(n, \lambda)$.

Example You are at a bank, and there are 3 people ahead of you. The serving time for each person is Exponential with mean 2 minutes. Only one person at a time can be served. The distribution of your waiting time until it's your turn to be served is $\text{Gamma}(3, \frac{1}{2})$.

Beta Distribution



Conjugate Prior of the Binomial In the Bayesian approach to statistics, parameters are viewed as random variables, to reflect our uncertainty. The *prior* for a parameter is its distribution before observing data. The *posterior* is the distribution for the parameter after observing data. Beta is the *conjugate* prior of the Binomial because if you have a Beta-distributed prior on p in a Binomial, then the posterior distribution on p given the Binomial data is also Beta-distributed. Consider the following two-level model:

$$X | p \sim \text{Bin}(n, p)$$
$$p \sim \text{Beta}(a, b)$$

Then after observing $X = x$, we get the posterior distribution

$$p | (X = x) \sim \text{Beta}(a + x, b + n - x)$$

Order statistics of the Uniform See *Order Statistics*.

Beta-Gamma relationship If $X \sim \text{Gamma}(a, \lambda)$, $Y \sim \text{Gamma}(b, \lambda)$, with $X \perp\!\!\!\perp Y$ then

- $\frac{X}{X+Y} \sim \text{Beta}(a, b)$
- $X + Y \perp\!\!\!\perp \frac{X}{X+Y}$

This is known as the **bank–post office result**.

χ^2 (Chi-Square) Distribution

Let us say that X is distributed χ^2_n. We know the following:

Story A Chi-Square(n) is the sum of the squares of n independent standard Normal r.v.s.

Properties and Representations

$$X \text{ is distributed as } Z_1^2 + Z_2^2 + \cdots + Z_n^2 \text{ for i.i.d. } Z_i \sim \mathcal{N}(0,1)$$

$$X \sim \text{Gamma}(n/2, 1/2)$$

Discrete Distributions

Distributions for four sampling schemes

	Replace	No Replace
Fixed # trials (n)	Binomial (Bern if $n = 1$)	HGeom
Draw until r success	NBin (Geom if $r = 1$)	NHGeom

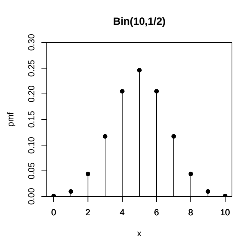
Bernoulli Distribution

The Bernoulli distribution is the simplest case of the Binomial distribution, where we only have one trial ($n = 1$). Let us say that X is distributed Bern(p). We know the following:

Story A trial is performed with probability p of "success", and X is the indicator of success: 1 means success, 0 means failure.

Example Let X be the indicator of Heads for a fair coin toss. Then $X \sim \text{Bern}(\frac{1}{2})$. Also, $1 - X \sim \text{Bern}(\frac{1}{2})$ is the indicator of Tails.

Binomial Distribution



Let us say that X is distributed Bin(n, p). We know the following:

Story X is the number of "successes" that we will achieve in n independent trials, where each trial is either a success or a failure, each with the same probability p of success. We can also write X as a sum of multiple independent Bern(p) random variables. Let $X \sim \text{Bin}(n, p)$ and $X_j \sim \text{Bern}(p)$, where all of the Bernoullis are independent. Then

$$X = X_1 + X_2 + X_3 + \cdots + X_n$$

Example If Jeremy Lin makes 10 free throws and each one independently has a $\frac{3}{4}$ chance of getting in, then the number of free throws he makes is distributed Bin($10, \frac{3}{4}$).

Properties Let $X \sim \text{Bin}(n, p), Y \sim \text{Bin}(m, p)$ with $X \perp\!\!\!\perp Y$.

- **Redefine success** $n - X \sim \text{Bin}(n, 1 - p)$
- **Sum** $X + Y \sim \text{Bin}(n + m, p)$
- **Conditional** $X|(X + Y = r) \sim \text{HGeom}(n, m, r)$
- **Binomial-Poisson Relationship** Bin(n, p) is approximately Pois(λ) if p is small.
- **Binomial-Normal Relationship** Bin(n, p) is approximately $\mathcal{N}(np, np(1 - p))$ if n is large and p is not near 0 or 1.

Geometric Distribution

Let us say that X is distributed Geom(p). We know the following:

Story X is the number of "failures" that we will achieve before we achieve our first success. Our successes have probability p.

Example If each pokeball we throw has probability $\frac{1}{10}$ to catch Mew, the number of failed pokeballs will be distributed Geom($\frac{1}{10}$).

First Success Distribution

Equivalent to the Geometric distribution, except that it includes the first success in the count. This is 1 more than the number of failures. If $X \sim \text{FS}(p)$ then $E(X) = 1/p$.

Negative Binomial Distribution

Let us say that X is distributed NBin(r, p). We know the following:

Story X is the number of "failures" that we will have before we achieve our rth success. Our successes have probability p.

Example Thundershock has 60% accuracy and can faint a wild Raticate in 3 hits. The number of misses before Pikachu faints Raticate with Thundershock is distributed NBin($3, 0.6$).

Hypergeometric Distribution

Let us say that X is distributed HGeom(w, b, n). We know the following:

Story In a population of w desired objects and b undesired objects, X is the number of "successes" we will have in a draw of n objects, without replacement. The draw of n objects is assumed to be a **simple random sample** (all sets of n objects are equally likely).

Examples Here are some HGeom examples.

- Let's say that we have only b Weedles (failure) and w Pikachus (success) in Viridian Forest. We encounter n Pokemon in the forest, and X is the number of Pikachus in our encounters.
- The number of Aces in a 5 card hand.
- You have w white balls and b black balls, and you draw n balls. You will draw X white balls.
- You have w white balls and b black balls, and you draw n balls without replacement. The number of white balls in your sample is HGeom(w, b, n); the number of black balls is HGeom(b, w, n).
- **Capture-recapture** A forest has N elk, you capture n of them, tag them, and release them. Then you recapture a new sample of size m. How many tagged elk are now in the new sample? HGeom($n, N - n, m$)

Poisson Distribution

Let us say that X is distributed Pois(λ). We know the following:

Story There are rare events (low probability events) that occur many different ways (high possibilities of occurences) at an average rate of λ occurrences per unit space or time. The number of events that occur in that unit of space or time is X.

Example A certain busy intersection has an average of 2 accidents per month. Since an accident is a low probability event that can happen many different ways, it is reasonable to model the number of accidents in a month at that intersection as Pois(2). Then the number of accidents that happen in two months at that intersection is distributed Pois(4).

Properties Let $X \sim \text{Pois}(\lambda_1)$ and $Y \sim \text{Pois}(\lambda_2)$, with $X \perp\!\!\!\perp Y$.

1. **Sum** $X + Y \sim \text{Pois}(\lambda_1 + \lambda_2)$
2. **Conditional** $X|(X + Y = n) \sim \text{Bin}\left(n, \frac{\lambda_1}{\lambda_1+\lambda_2}\right)$
3. **Chicken-egg** If there are $Z \sim \text{Pois}(\lambda)$ items and we randomly and independently "accept" each item with probability p, then the number of accepted items $Z_1 \sim \text{Pois}(\lambda p)$, and the number of rejected items $Z_2 \sim \text{Pois}(\lambda(1 - p))$, and $Z_1 \perp\!\!\!\perp Z_2$.

Multivariate Distributions

Multinomial Distribution

Let us say that the vector $\vec{X} = (X_1, X_2, X_3, \ldots, X_k) \sim \text{Mult}_k(n, \vec{p})$ where $\vec{p} = (p_1, p_2, \ldots, p_k)$.

Story We have n items, which can fall into any one of the k buckets independently with the probabilities $\vec{p} = (p_1, p_2, \ldots, p_k)$.

Example Let us assume that every year, 100 students in the Harry Potter Universe are randomly and independently sorted into one of four houses with equal probability. The number of people in each of the houses is distributed $\text{Mult}_4(100, \vec{p})$, where $\vec{p} = (0.25, 0.25, 0.25, 0.25)$. Note that $X_1 + X_2 + \cdots + X_4 = 100$, and they are dependent.

Joint PMF For $n = n_1 + n_2 + \cdots + n_k$,

$$P(\vec{X} = \vec{n}) = \frac{n!}{n_1!n_2!\ldots n_k!} p_1^{n_1} p_2^{n_2} \ldots p_k^{n_k}$$

Marginal PMF, Lumping, and Conditionals Marginally, $X_i \sim \text{Bin}(n, p_i)$ since we can define "success" to mean category i. If you lump together multiple categories in a Multinomial, then it is still Multinomial. For example, $X_i + X_j \sim \text{Bin}(n, p_i + p_j)$ for $i \neq j$ since we can define "success" to mean being in category i or j. Similarly, if $k = 6$ and we lump categories 1-2 and lump categories 3-5, then

$$(X_1 + X_2, X_3 + X_4 + X_5, X_6) \sim \text{Mult}_3(n, (p_1 + p_2, p_3 + p_4 + p_5, p_6))$$

Conditioning on some X_j also still gives a Multinomial:

$$X_1, \ldots, X_{k-1}|X_k = n_k \sim \text{Mult}_{k-1}\left(n - n_k, \left(\frac{p_1}{1 - p_k}, \ldots, \frac{p_{k-1}}{1 - p_k}\right)\right)$$

Variances and Covariances We have $X_i \sim \text{Bin}(n, p_i)$ marginally, so $\text{Var}(X_i) = np_i(1 - p_i)$. Also, $\text{Cov}(X_i, X_j) = -np_ip_j$ for $i \neq j$.

Multivariate Uniform Distribution

See the univariate Uniform for stories and examples. For the 2D Uniform on some region, probability is proportional to area. Every point in the support has equal density, of value $\frac{1}{\text{area of region}}$. For the 3D Uniform, probability is proportional to volume.

Multivariate Normal (MVN) Distribution

A vector $\vec{X} = (X_1, X_2, \ldots, X_k)$ is Multivariate Normal if every linear combination is Normally distributed, i.e., $t_1X_1 + t_2X_2 + \cdots + t_kX_k$ is Normal for any constants $t_1, t_2, \ldots, t_k$. The parameters of the Multivariate Normal are the **mean vector** $\vec{\mu} = (\mu_1, \mu_2, \ldots, \mu_k)$ and the **covariance matrix** where the (i, j) entry is $\text{Cov}(X_i, X_j)$.

Properties The Multivariate Normal has the following properties.

- Any subvector is also MVN.
- If any two elements within an MVN are uncorrelated, then they are independent.
- The joint PDF of a Bivariate Normal (X, Y) with $\mathcal{N}(0, 1)$ marginal distributions and correlation $\rho \in (-1, 1)$ is

$$f_{X,Y}(x, y) = \frac{1}{2\pi\tau} \exp\left(-\frac{1}{2\tau^2}(x^2 + y^2 - 2\rho xy)\right),$$

with $\tau = \sqrt{1 - \rho^2}$.

Distribution Properties

Important CDFs

Standard Normal Φ

Exponential(λ) $F(x) = 1 - e^{-\lambda x}$, for $x \in (0, \infty)$

Uniform(0,1) $F(x) = x$, for $x \in (0,1)$

Convolutions of Random Variables

A convolution of n random variables is simply their sum. For the following results, let X and Y be *independent*.

1. $X \sim \text{Pois}(\lambda_1)$, $Y \sim \text{Pois}(\lambda_2) \longrightarrow X + Y \sim \text{Pois}(\lambda_1 + \lambda_2)$
2. $X \sim \text{Bin}(n_1, p)$, $Y \sim \text{Bin}(n_2, p) \longrightarrow X + Y \sim \text{Bin}(n_1 + n_2, p)$. $\text{Bin}(n, p)$ can be thought of as a sum of i.i.d. $\text{Bern}(p)$ r.v.s.
3. $X \sim \text{Gamma}(a_1, \lambda)$, $Y \sim \text{Gamma}(a_2, \lambda)$ $\longrightarrow X + Y \sim \text{Gamma}(a_1 + a_2, \lambda)$. $\text{Gamma}(n, \lambda)$ with n an integer can be thought of as a sum of i.i.d. $\text{Expo}(\lambda)$ r.v.s.
4. $X \sim \text{NBin}(r_1, p)$, $Y \sim \text{NBin}(r_2, p)$ $\longrightarrow X + Y \sim \text{NBin}(r_1 + r_2, p)$. $\text{NBin}(r, p)$ can be thought of as a sum of i.i.d. $\text{Geom}(p)$ r.v.s.
5. $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$, $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$ $\longrightarrow X + Y \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

Special Cases of Distributions

1. $\text{Bin}(1, p) \sim \text{Bern}(p)$
2. $\text{Beta}(1, 1) \sim \text{Unif}(0, 1)$
3. $\text{Gamma}(1, \lambda) \sim \text{Expo}(\lambda)$
4. $\chi_n^2 \sim \text{Gamma}\left(\frac{n}{2}, \frac{1}{2}\right)$
5. $\text{NBin}(1, p) \sim \text{Geom}(p)$

Inequalities

1. **Cauchy-Schwarz** $|E(XY)| \leq \sqrt{E(X^2)E(Y^2)}$
2. **Markov** $P(X \geq a) \leq \frac{E|X|}{a}$ for $a > 0$
3. **Chebyshev** $P(|X - \mu| \geq a) \leq \frac{\sigma^2}{a^2}$ for $E(X) = \mu, \text{Var}(X) = \sigma^2$
4. **Jensen** $E(g(X)) \geq g(E(X))$ for g convex; reverse if g is concave

Formulas

Geometric Series

$$1 + r + r^2 + \cdots + r^{n-1} = \sum_{k=0}^{n-1} r^k = \frac{1 - r^n}{1 - r}$$

$$1 + r + r^2 + \cdots = \frac{1}{1 - r} \text{ if } |r| < 1$$

Exponential Function (e^x)

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!} = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots = \lim_{n \to \infty} \left(1 + \frac{x}{n}\right)^n$$

Gamma and Beta Integrals

You can sometimes solve complicated-looking integrals by pattern-matching to a gamma or beta integral:

$$\int_0^\infty x^{t-1} e^{-x}\, dx = \Gamma(t) \qquad \int_0^1 x^{a-1}(1-x)^{b-1}\, dx = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

Also, $\Gamma(a + 1) = a\Gamma(a)$, and $\Gamma(n) = (n - 1)!$ if n is a positive integer.

Euler's Approximation for Harmonic Sums

$$1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n} \approx \log n + 0.577\ldots$$

Stirling's Approximation for Factorials

$$n! \approx \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$$

Miscellaneous Definitions

Medians and Quantiles Let X have CDF F. Then X has median m if $F(m) \geq 0.5$ and $P(X \geq m) \geq 0.5$. For X continuous, m satisfies $F(m) = 1/2$. In general, the ath quantile of X is $\min\{x : F(x) \geq a\}$; the median is the case $a = 1/2$.

log Statisticians generally use log to refer to natural log (i.e., base e).

i.i.d r.v.s Independent, identically-distributed random variables.

Example Problems

Contributions from Sebastian Chiu

Calculating Probability

A textbook has n typos, which are randomly scattered amongst its n pages, independently. You pick a random page. What is the probability that it has no typos? **Answer:** There is a $\left(1 - \frac{1}{n}\right)$ probability that any specific typo isn't on your page, and thus a $\boxed{\left(1 - \frac{1}{n}\right)^n}$ probability that there are no typos on your page. For n large, this is approximately $e^{-1} = 1/e$.

Linearity and Indicators (1)

In a group of n people, what is the expected number of distinct birthdays (month and day)? What is the expected number of birthday matches? **Answer:** Let X be the number of distinct birthdays and I_j be the indicator for the jth day being represented.

$$E(I_j) = 1 - P(\text{no one born on day } j) = 1 - (364/365)^n$$

By linearity, $\boxed{E(X) = 365\left(1 - (364/365)^n\right)}$. Now let Y be the number of birthday matches and J_i be the indicator that the ith pair of people have the same birthday. The probability that any two specific people share a birthday is $1/365$, so $\boxed{E(Y) = \binom{n}{2}/365}$.

Linearity and Indicators (2)

This problem is commonly known as the ***hat-matching problem****.* There are n people at a party, each with hat. At the end of the party, they each leave with a random hat. What is the expected number of people who leave with the right hat? **Answer:** Each hat has a $1/n$ chance of going to the right person. By linearity, the average number of hats that go to their owners is $\boxed{n(1/n) = 1}$.

Linearity and First Success

This problem is commonly known as the ***coupon collector problem****.* There are n coupon types. At each draw, you get a uniformly random coupon type. What is the expected number of coupons needed until you have a complete set? **Answer:** Let N be the number of coupons needed; we want $E(N)$. Let $N = N_1 + \cdots + N_n$, where N_1 is the draws to get our first new coupon, N_2 is the *additional* draws needed to draw our second new coupon and so on. By the story of the First Success, $N_2 \sim \text{FS}((n-1)/n)$ (after collecting first coupon type, there's $(n-1)/n$ chance you'll get something new). Similarly, $N_3 \sim \text{FS}((n-2)/n)$, and $N_j \sim \text{FS}((n-j+1)/n)$. By linearity,

$$E(N) = E(N_1) + \cdots + E(N_n) = \frac{n}{n} + \frac{n}{n-1} + \cdots + \frac{n}{1} = \boxed{n\sum_{j=1}^{n}\frac{1}{j}}$$

This is approximately $n(\log(n) + 0.577)$ by Euler's approximation.

Orderings of i.i.d. random variables

I call 2 UberX's and 3 Lyfts at the same time. If the time it takes for the rides to reach me are i.i.d., what is the probability that all the Lyfts will arrive first? **Answer:** Since the arrival times of the five cars are i.i.d., all 5! orderings of the arrivals are equally likely. There are 3!2! orderings that involve the Lyfts arriving first, so the probability that the Lyfts arrive first is $\boxed{\frac{3!2!}{5!} = 1/10}$. Alternatively, there are $\binom{5}{3}$ ways to choose 3 of the 5 slots for the Lyfts to occupy, where each of the choices are equally likely. One of these choices has all 3 of the Lyfts arriving first, so the probability is $\boxed{1/\binom{5}{3} = 1/10}$.

Expectation of Negative Hypergeometric

What is the expected number of cards that you draw before you pick your first Ace in a shuffled deck (not counting the Ace)? **Answer:** Consider a non-Ace. Denote this to be card j. Let I_j be the indicator that card j will be drawn before the first Ace. Note that $I_j = 1$ says that j is before all 4 of the Aces in the deck. The probability that this occurs is $1/5$ by symmetry. Let X be the number of cards drawn before the first Ace. Then $X = I_1 + I_2 + \ldots + I_{48}$, where each indicator corresponds to one of the 48 non-Aces. Thus,

$$E(X) = E(I_1) + E(I_2) + \ldots + E(I_{48}) = 48/5 = \boxed{9.6}.$$

Minimum and Maximum of RVs

What is the CDF of the maximum of n independent Unif(0,1) random variables? **Answer:** Note that for r.v.s $X_1, X_2, \ldots, X_n$,

$$P(\min(X_1, X_2, \ldots, X_n) \geq a) = P(X_1 \geq a, X_2 \geq a, \ldots, X_n \geq a)$$

Similarly,

$$P(\max(X_1, X_2, \ldots, X_n) \leq a) = P(X_1 \leq a, X_2 \leq a, \ldots, X_n \leq a)$$

We will use this principle to find the CDF of $U_{(n)}$, where $U_{(n)} = \max(U_1, U_2, \ldots, U_n)$ and $U_i \sim \text{Unif}(0, 1)$ are i.i.d.

$$\begin{aligned} P(\max(U_1, U_2, \ldots, U_n) \leq a) &= P(U_1 \leq a, U_2 \leq a, \ldots, U_n \leq a) \\ &= P(U_1 \leq a)P(U_2 \leq a) \ldots P(U_n \leq a) \\ &= \boxed{a^n} \end{aligned}$$

for $0 < a < 1$ (and the CDF is 0 for $a \leq 0$ and 1 for $a \geq 1$).

Pattern-matching with e^x Taylor series

For $X \sim \text{Pois}(\lambda)$, find $E\left(\frac{1}{X+1}\right)$. **Answer:** By LOTUS,

$$E\left(\frac{1}{X+1}\right) = \sum_{k=0}^{\infty} \frac{1}{k+1}\frac{e^{-\lambda}\lambda^k}{k!} = \frac{e^{-\lambda}}{\lambda}\sum_{k=0}^{\infty}\frac{\lambda^{k+1}}{(k+1)!} = \boxed{\frac{e^{-\lambda}}{\lambda}(e^\lambda - 1)}$$

Adam's Law and Eve's Law

William really likes speedsolving Rubik's Cubes. But he's pretty bad at it, so sometimes he fails. On any given day, William will attempt $N \sim \text{Geom}(s)$ Rubik's Cubes. Suppose each time, he has probability p of solving the cube, independently. Let T be the number of Rubik's Cubes he solves during a day. Find the mean and variance of T.
Answer: Note that $T|N \sim \text{Bin}(N, p)$. So by Adam's Law,

$$E(T) = E(E(T|N)) = E(Np) = \boxed{\frac{p(1-s)}{s}}$$

Similarly, by Eve's Law, we have that

$$\text{Var}(T) = E(\text{Var}(T|N)) + \text{Var}(E(T|N)) = E(Np(1-p)) + \text{Var}(Np)$$
$$= \frac{p(1-p)(1-s)}{s} + \frac{p^2(1-s)}{s^2} = \boxed{\frac{p(1-s)(p+s(1-p))}{s^2}}$$

MGF – Finding Moments

Find $E(X^3)$ for $X \sim \text{Expo}(\lambda)$ using the MGF of X. **Answer:** The MGF of an Expo(λ) is $M(t) = \frac{\lambda}{\lambda - t}$. To get the third moment, we can take the third derivative of the MGF and evaluate at $t = 0$:

$$\boxed{E(X^3) = \frac{6}{\lambda^3}}$$

But a much nicer way to use the MGF here is via pattern recognition: note that $M(t)$ looks like it came from a geometric series:

$$\frac{1}{1 - \frac{t}{\lambda}} = \sum_{n=0}^{\infty} \left(\frac{t}{\lambda}\right)^n = \sum_{n=0}^{\infty} \frac{n!}{\lambda^n} \frac{t^n}{n!}$$

The coefficient of $\frac{t^n}{n!}$ here is the nth moment of X, so we have $E(X^n) = \frac{n!}{\lambda^n}$ for all nonnegative integers n.

Markov chains (1)

Suppose X_n is a two-state Markov chain with transition matrix

$$Q = \begin{array}{c|cc} & 0 & 1 \\ 0 & 1-\alpha & \alpha \\ 1 & \beta & 1-\beta \end{array}$$

Find the stationary distribution $\vec{s} = (s_0, s_1)$ of X_n by solving $\vec{s}Q = \vec{s}$, and show that the chain is reversible with respect to $\vec{s}$. **Answer:** The equation $\vec{s}Q = \vec{s}$ says that

$$s_0 = s_0(1-\alpha) + s_1\beta \text{ and } s_1 = s_0(\alpha) + s_0(1-\beta)$$

By solving this system of linear equations, we have

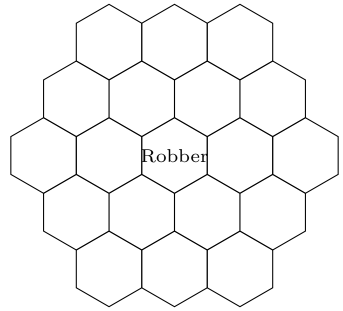
$$\boxed{\vec{s} = \left(\frac{\beta}{\alpha+\beta}, \frac{\alpha}{\alpha+\beta}\right)}$$

To show that the chain is reversible with respect to $\vec{s}$, we must show $s_i q_{ij} = s_j q_{ji}$ for all i, j. This is done if we can show $s_0 q_{01} = s_1 q_{10}$. And indeed,

$$s_0 q_{01} = \frac{\alpha\beta}{\alpha+\beta} = s_1 q_{10}$$

Markov chains (2)

William and Sebastian play a modified game of Settlers of Catan, where every turn they randomly move the robber (which starts on the center tile) to one of the adjacent hexagons.



(a) Is this Markov chain irreducible? Is it aperiodic? **Answer:** $\boxed{\text{Yes to both.}}$ The Markov chain is irreducible because it can get from anywhere to anywhere else. The Markov chain is aperiodic because the robber can return back to a square in $2, 3, 4, 5, \ldots$ moves, and the GCD of those numbers is 1.

(b) What is the stationary distribution of this Markov chain? **Answer:** Since this is a random walk on an undirected graph, the stationary distribution is proportional to the degree sequence. The degree for the corner pieces is 3, the degree for the edge pieces is 4, and the degree for the center pieces is 6. To normalize this degree sequence, we divide by its sum. The sum of the degrees is $6(3) + 6(4) + 7(6) = 84$. Thus the stationary probability of being on a corner is $3/84 = 1/28$, on an edge is $4/84 = 1/21$, and in the center is $6/84 = 1/14$.

(c) What fraction of the time will the robber be in the center tile in this game, in the long run? **Answer:** By the above, $\boxed{1/14}$.

(d) What is the expected amount of moves it will take for the robber to return to the center tile? **Answer:** Since this chain is irreducible and aperiodic, to get the expected time to return we can just invert the stationary probability. Thus on average it will take $\boxed{14}$ turns for the robber to return to the center tile.

Problem-Solving Strategies

Contributions from Jessy Hwang, Yuan Jiang, Yuqi Hou

1. **Getting started.** Start by *defining relevant events and random variables.* ("Let A be the event that I pick the fair coin"; "Let X be the number of successes.") Clear notion is important for clear thinking! Then decide what it is that you're supposed to be finding, in terms of your notation ("I want to find $P(X = 3|A)$"). Think about what type of object your answer should be (a number? A random variable? A PMF? A PDF?) and what it should be in terms of.

   Try simple and extreme cases. To make an abstract experiment more concrete, try *drawing a picture* or making up numbers that could have happened. Pattern recognition: does the structure of the problem resemble something we've seen before?

2. **Calculating probability of an event.** Use counting principles if the naive definition of probability applies. Is the probability of the complement easier to find? Look for symmetries. Look for something to condition on, then apply Bayes' Rule or the Law of Total Probability.

3. **Finding the distribution of a random variable.** First make sure you need the full distribution not just the mean (see next item). Check the *support* of the random variable: what values can it take on? Use this to rule out distributions that don't fit. Is there a *story* for one of the named distributions that fits the problem at hand? Can you write the random variable as a function of an r.v. with a known distribution, say $Y = g(X)$?

4. **Calculating expectation.** If it has a named distribution, check out the table of distributions. If it's a function of an r.v. with a named distribution, try LOTUS. If it's a count of something, try breaking it up into indicator r.v.s. If you can condition on something natural, consider using Adam's law.

5. **Calculating variance.** Consider independence, named distributions, and LOTUS. If it's a count of something, break it up into a sum of indicator r.v.s. If it's a sum, use properties of covariance. If you can condition on something natural, consider using Eve's Law.

6. **Calculating $E(X^2)$.** Do you already know $E(X)$ or $\text{Var}(X)$? Recall that $\text{Var}(X) = E(X^2) - (E(X))^2$. Otherwise try LOTUS.

7. **Calculating covariance.** Use the properties of covariance. If you're trying to find the covariance between two components of a Multinomial distribution, X_i, X_j, then the covariance is $-np_ip_j$ for $i \neq j$.

8. **Symmetry.** If $X_1, \ldots, X_n$ are i.i.d., consider using symmetry.

9. **Calculating probabilities of orderings.** Remember that all $n!$ ordering of i.i.d. continuous random variables $X_1, \ldots, X_n$ are equally likely.

10. **Determining independence.** There are several equivalent definitions. Think about simple and extreme cases to see if you can find a counterexample.

11. **Do a painful integral.** If your integral looks painful, see if you can write your integral in terms of a known PDF (like Gamma or Beta), and use the fact that PDFs integrate to 1?

12. **Before moving on.** Check some simple and extreme cases, check whether the answer seems plausible, check for biohazards.

Biohazards

Contributions from Jessy Hwang

1. **Don't misuse the naive definition of probability.** When answering "What is the probability that in a group of 3 people, no two have the same birth month?", it is *not* correct to treat the people as indistinguishable balls being placed into 12 boxes, since that assumes the list of birth months {January, January, January} is just as likely as the list {January, April, June}, even though the latter is six times more likely.

2. **Don't confuse unconditional, conditional, and joint probabilities.** In applying $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$, it is *not* correct to say "$P(B) = 1$ because we know B happened"; $P(B)$ is the *prior* probability of B. Don't confuse $P(A|B)$ with $P(A, B)$.

3. **Don't assume independence without justification.** In the matching problem, the probability that card 1 is a match and card 2 is a match is not $1/n^2$. Binomial and Hypergeometric are often confused; the trials are independent in the Binomial story and dependent in the Hypergeometric story.

4. **Don't forget to do sanity checks.** Probabilities must be between 0 and 1. Variances must be ≥ 0. Supports must make sense. PMFs must sum to 1. PDFs must integrate to 1.

5. **Don't confuse random variables, numbers, and events.** Let X be an r.v. Then $g(X)$ is an r.v. for any function g. In particular, X^2, $|X|$, $F(X)$, and $I_{X>3}$ are r.v.s. $P(X^2 < X | X \geq 0)$, $E(X)$, $\text{Var}(X)$, and $g(E(X))$ are numbers. $X = 2$ and $F(X) \geq -1$ are events. It does not make sense to write $\int_{-\infty}^{\infty} F(X)dx$, because $F(X)$ is a random variable. It does not make sense to write $P(X)$, because X is not an event.

6. **Don't confuse a random variable with its distribution.** To get the PDF of X^2, you can't just square the PDF of X. The right way is to use transformations. To get the PDF of $X + Y$, you can't just add the PDF of X and the PDF of Y. The right way is to compute the convolution.

7. **Don't pull non-linear functions out of expectations.** $E(g(X))$ does not equal $g(E(X))$ in general. The St. Petersburg paradox is an extreme example. See also Jensen's inequality. The right way to find $E(g(X))$ is with LOTUS.

Distributions in R

Command	What it does
`help(distributions)`	shows documentation on distributions
`dbinom(k,n,p)`	PMF $P(X = k)$ for $X \sim \text{Bin}(n, p)$
`pbinom(x,n,p)`	CDF $P(X \leq x)$ for $X \sim \text{Bin}(n, p)$
`qbinom(a,n,p)`	ath quantile for $X \sim \text{Bin}(n, p)$
`rbinom(r,n,p)`	vector of r i.i.d. $\text{Bin}(n, p)$ r.v.s
`dgeom(k,p)`	PMF $P(X = k)$ for $X \sim \text{Geom}(p)$
`dhyper(k,w,b,n)`	PMF $P(X = k)$ for $X \sim \text{HGeom}(w, b, n)$
`dnbinom(k,r,p)`	PMF $P(X = k)$ for $X \sim \text{NBin}(r, p)$
`dpois(k,r)`	PMF $P(X = k)$ for $X \sim \text{Pois}(r)$
`dbeta(x,a,b)`	PDF $f(x)$ for $X \sim \text{Beta}(a, b)$
`dchisq(x,n)`	PDF $f(x)$ for $X \sim \chi^2_n$
`dexp(x,b)`	PDF $f(x)$ for $X \sim \text{Expo}(b)$
`dgamma(x,a,r)`	PDF $f(x)$ for $X \sim \text{Gamma}(a, r)$
`dlnorm(x,m,s)`	PDF $f(x)$ for $X \sim \mathcal{LN}(m, s^2)$
`dnorm(x,m,s)`	PDF $f(x)$ for $X \sim \mathcal{N}(m, s^2)$
`dt(x,n)`	PDF $f(x)$ for $X \sim t_n$
`dunif(x,a,b)`	PDF $f(x)$ for $X \sim \text{Unif}(a, b)$

The table above gives R commands for working with various named distributions. Commands analogous to `pbinom`, `qbinom`, and `rbinom` work for the other distributions in the table. For example, `pnorm`, `qnorm`, and `rnorm` can be used to get the CDF, quantiles, and random generation for the Normal. For the Multinomial, `dmultinom` can be used for calculating the joint PMF and `rmultinom` can be used for generating random vectors. For the Multivariate Normal, after installing and loading the `mvtnorm` package `dmvnorm` can be used for calculating the joint PDF and `rmvnorm` can be used for generating random vectors.

Recommended Resources

- Introduction to Probability Book (`http://bit.ly/introprobability`)
- Stat 110 Online (`http://stat110.net`)
- Stat 110 Quora Blog (`https://stat110.quora.com/`)
- Quora Probability FAQ (`http://bit.ly/probabilityfaq`)
- R Studio (`https://www.rstudio.com`)
- LaTeX File (`github.com/wzchen/probability_cheatsheet`)

Please share this cheatsheet with friends!
`http://wzchen.com/probability-cheatsheet`

Table of Distributions

Distribution	PMF/PDF and Support	Expected Value	Variance	MGF
Bernoulli Bern(p)	$P(X=1)=p$ $P(X=0)=q=1-p$	p	pq	$q+pe^t$
Binomial Bin(n,p)	$P(X=k)=\binom{n}{k}p^k q^{n-k}$ $k\in\{0,1,2,\dots n\}$	np	npq	$(q+pe^t)^n$
Geometric Geom(p)	$P(X=k)=q^k p$ $k\in\{0,\,1,\,2,\,\dots\}$	q/p	q/p^2	$\frac{p}{1-qe^t}$, $qe^t<1$
Negative Binomial NBin(r,p)	$P(X=n)=\binom{r+n-1}{r-1}p^r q^n$ $n\in\{0,\,1,\,2,\,\dots\}$	rq/p	rq/p^2	$(\frac{p}{1-qe^t})^r$, $qe^t<1$
Hypergeometric HGeom(w,b,n)	$P(X=k)=\binom{w}{k}\binom{b}{n-k}/\binom{w+b}{n}$ $k\in\{0,1,2,\dots,n\}$	$\mu=\frac{nw}{b+w}$	$\left(\frac{w+b-n}{w+b-1}\right)n\frac{\mu}{n}(1-\frac{\mu}{n})$	messy
Poisson Pois(λ)	$P(X=k)=\frac{e^{-\lambda}\lambda^k}{k!}$ $k\in\{0,\,1,\,2,\,\dots\}$	λ	λ	$e^{\lambda(e^t-1)}$
Uniform Unif(a,b)	$f(x)=\frac{1}{b-a}$ $x\in(a,b)$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	$\frac{e^{tb}-e^{ta}}{t(b-a)}$
Normal $\mathcal{N}(\mu,\sigma^2)$	$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-(x\,-\,\mu)^2/(2\sigma^2)}$ $x\in(-\infty,\infty)$	μ	σ^2	$e^{t\mu+\frac{\sigma^2t^2}{2}}$
Exponential Expo(λ)	$f(x)=\lambda e^{-\lambda x}$ $x\in(0,\infty)$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	$\frac{\lambda}{\lambda-t}$, $t<\lambda$
Gamma Gamma(a,λ)	$f(x)=\frac{1}{\Gamma(a)}(\lambda x)^a e^{-\lambda x}\frac{1}{x}$ $x\in(0,\infty)$	$\frac{a}{\lambda}$	$\frac{a}{\lambda^2}$	$\left(\frac{\lambda}{\lambda-t}\right)^a$, $t<\lambda$
Beta Beta(a,b)	$f(x)=\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}x^{a-1}(1-x)^{b-1}$ $x\in(0,1)$	$\mu=\frac{a}{a+b}$	$\frac{\mu(1-\mu)}{(a+b+1)}$	messy
Log-Normal $\mathcal{LN}(\mu,\sigma^2)$	$\frac{1}{x\sigma\sqrt{2\pi}}e^{-(\log x-\mu)^2/(2\sigma^2)}$ $x\in(0,\infty)$	$\theta=e^{\mu+\sigma^2/2}$	$\theta^2(e^{\sigma^2}-1)$	doesn't exist
Chi-Square χ^2_n	$\frac{1}{2^{n/2}\Gamma(n/2)}x^{n/2-1}e^{-x/2}$ $x\in(0,\infty)$	n	$2n$	$(1-2t)^{-n/2}$, $t<1/2$
Student-t t_n	$\frac{\Gamma((n+1)/2)}{\sqrt{n\pi}\Gamma(n/2)}(1+x^2/n)^{-(n+1)/2}$ $x\in(-\infty,\infty)$	0 if $n>1$	$\frac{n}{n-2}$ if $n>2$	doesn't exist

Chapter 13

PROBABILITY

❖ *The theory of probabilities is simply the Science of logic quantitatively treated. – C.S. PEIRCE* ❖

13.1 Introduction

Pierre de Fermat
(1601-1665)

In earlier Classes, we have studied the probability as a measure of uncertainty of events in a random experiment. We discussed the axiomatic approach formulated by Russian Mathematician, A.N. Kolmogorov (1903-1987) and treated probability as a function of outcomes of the experiment. We have also established equivalence between the axiomatic theory and the classical theory of probability in case of equally likely outcomes. On the basis of this relationship, we obtained probabilities of events associated with discrete sample spaces. We have also studied the addition rule of probability. In this chapter, we shall discuss the important concept of conditional probability of an event given that another event has occurred, which will be helpful in understanding the Bayes' theorem, multiplication rule of probability and independence of events. We shall also learn an important concept of random variable and its probability distribution and also the mean and variance of a probability distribution. In the last section of the chapter, we shall study an important discrete probability distribution called Binomial distribution. Throughout this chapter, we shall take up the experiments having equally likely outcomes, unless stated otherwise.

13.2 Conditional Probability

Uptill now in probability, we have discussed the methods of finding the probability of events. If we have two events from the same sample space, does the information about the occurrence of one of the events affect the probability of the other event? Let us try to answer this question by taking up a random experiment in which the outcomes are equally likely to occur.

Consider the experiment of tossing three fair coins. The sample space of the experiment is

$$S = \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}$$

Since the coins are fair, we can assign the probability $\frac{1}{8}$ to each sample point. Let E be the event 'at least two heads appear' and F be the event 'first coin shows tail'. Then

$$E = \{HHH, HHT, HTH, THH\}$$

and $$F = \{THH, THT, TTH, TTT\}$$

Therefore $$P(E) = P(\{HHH\}) + P(\{HHT\}) + P(\{HTH\}) + P(\{THH\})$$

$$= \frac{1}{8}+\frac{1}{8}+\frac{1}{8}+\frac{1}{8}=\frac{1}{2} \text{ (Why ?)}$$

and $$P(F) = P(\{THH\}) + P(\{THT\}) + P(\{TTH\}) + P(\{TTT\})$$

$$= \frac{1}{8}+\frac{1}{8}+\frac{1}{8}+\frac{1}{8}=\frac{1}{2}$$

Also $$E \cap F = \{THH\}$$

with $$P(E \cap F) = P(\{THH\}) = \frac{1}{8}$$

Now, suppose we are given that the first coin shows tail, i.e. F occurs, then what is the probability of occurrence of E? With the information of occurrence of F, we are sure that the cases in which first coin does not result into a tail should not be considered while finding the probability of E. This information reduces our sample space from the set S to its subset F for the event E. In other words, the additional information really amounts to telling us that the situation may be considered as being that of a new random experiment for which the sample space consists of all those outcomes only which are favourable to the occurrence of the event F.

Now, the sample point of F which is favourable to event E is THH.

Thus, Probability of E considering F as the sample space = $\frac{1}{4}$,

or Probability of E given that the event F has occurred = $\frac{1}{4}$

This probability of the event E is called the *conditional probability of E given that F has already occurred*, and is denoted by P (E|F).

Thus $$P(E|F) = \frac{1}{4}$$

Note that the elements of F which favour the event E are the common elements of E and F, i.e. the sample points of $E \cap F$.

Thus, we can also write the conditional probability of E given that F has occurred as

$$P(E|F) = \frac{\text{Number of elementary events favourable to } E \cap F}{\text{Number of elementary events which are favourable to } F}$$

$$= \frac{n(E \cap F)}{n(F)}$$

Dividing the numerator and the denominator by total number of elementary events of the sample space, we see that P(E|F) can also be written as

$$P(E|F) = \frac{\dfrac{n(E \cap F)}{n(S)}}{\dfrac{n(F)}{n(S)}} = \frac{P(E \cap F)}{P(F)} \qquad \ldots (1)$$

Note that (1) is valid only when $P(F) \neq 0$ i.e., $F \neq \phi$ (Why?)

Thus, we can define the conditional probability as follows :

Definition 1 If E and F are two events associated with the same sample space of a random experiment, the conditional probability of the event E given that F has occurred, i.e. P (E|F) is given by

$$P(E|F) = \frac{P(E \cap F)}{P(F)} \text{ provided } P(F) \neq 0$$

13.2.1 *Properties of conditional probability*

Let E and F be events of a sample space S of an experiment, then we have

Property 1 $P(S|F) = P(F|F) = 1$

We know that

$$P(S|F) = \frac{P(S \cap F)}{P(F)} = \frac{P(F)}{P(F)} = 1$$

Also $$P(F|F) = \frac{P(F \cap F)}{P(F)} = \frac{P(F)}{P(F)} = 1$$

Thus $$P(S|F) = P(F|F) = 1$$

Property 2 *If* A *and* B *are any two events of a sample space* S *and* F *is an event of* S *such that* $P(F) \neq 0$, *then*

$$P((A \cup B)|F) = P(A|F) + P(B|F) - P((A \cap B)|F)$$

In particular, if A *and* B *are disjoint events, then*

$$P((A \cup B)|F) = P(A|F) + P(B|F)$$

We have

$$P((A \cup B)|F) = \frac{P[(A \cup B) \cap F]}{P(F)}$$

$$= \frac{P[(A \cap F) \cup (B \cap F)]}{P(F)}$$

(by distributive law of union of sets over intersection)

$$= \frac{P(A \cap F) + P(B \cap F) - P(A \cap B \cap F)}{P(F)}$$

$$= \frac{P(A \cap F)}{P(F)} + \frac{P(B \cap F)}{P(F)} - \frac{P[(A \cap B) \cap F]}{P(F)}$$

$$= P(A|F) + P(B|F) - P((A \cap B)|F)$$

When A and B are disjoint events, then

$$P((A \cap B)|F) = 0$$

$$\Rightarrow \quad P((A \cup B)|F) = P(A|F) + P(B|F)$$

Property 3 $P(E'|F) = 1 - P(E|F)$

From Property 1, we know that $P(S|F) = 1$

$\Rightarrow \quad P(E \cup E'|F) = 1$ since $S = E \cup E'$

$\Rightarrow \quad P(E|F) + P(E'|F) = 1$ since E and E′ are disjoint events

Thus, $\quad P(E'|F) = 1 - P(E|F)$

Let us now take up some examples.

Example 1 If $P(A) = \frac{7}{13}$, $P(B) = \frac{9}{13}$ and $P(A \cap B) = \frac{4}{13}$, evaluate $P(A|B)$.

Solution We have $P(A|B) = \dfrac{P(A \cap B)}{P(B)} = \dfrac{\frac{4}{13}}{\frac{9}{13}} = \dfrac{4}{9}$

Example 2 A family has two children. What is the probability that both the children are boys given that at least one of them is a boy ?

Solution Let b stand for boy and g for girl. The sample space of the experiment is

$$S = \{(b, b), (g, b), (b, g), (g, g)\}$$

Let E and F denote the following events :

E : 'both the children are boys'

F : 'at least one of the child is a boy'

Then $$E = \{(b,b)\} \text{ and } F = \{(b,b), (g,b), (b,g)\}$$

Now $$E \cap F = \{(b,b)\}$$

Thus $$P(F) = \frac{3}{4} \text{ and } P(E \cap F) = \frac{1}{4}$$

Therefore $$P(E|F) = \frac{P(E \cap F)}{P(F)} = \frac{\frac{1}{4}}{\frac{3}{4}} = \frac{1}{3}$$

Example 3 Ten cards numbered 1 to 10 are placed in a box, mixed up thoroughly and then one card is drawn randomly. If it is known that the number on the drawn card is more than 3, what is the probability that it is an even number?

Solution Let A be the event 'the number on the card drawn is even' and B be the event 'the number on the card drawn is greater than 3'. We have to find P(A|B).

Now, the sample space of the experiment is S = {1, 2, 3, 4, 5, 6, 7, 8, 9, 10}

Then $$A = \{2, 4, 6, 8, 10\}, \; B = \{4, 5, 6, 7, 8, 9, 10\}$$

and $$A \cap B = \{4, 6, 8, 10\}$$

Also $$P(A) = \frac{5}{10}, \; P(B) = \frac{7}{10} \text{ and } P(A \cap B) = \frac{4}{10}$$

Then $$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{4}{10}}{\frac{7}{10}} = \frac{4}{7}$$

Example 4 In a school, there are 1000 students, out of which 430 are girls. It is known that out of 430, 10% of the girls study in class XII. What is the probability that a student chosen randomly studies in Class XII given that the chosen student is a girl?

Solution Let E denote the event that a student chosen randomly studies in Class XII and F be the event that the randomly chosen student is a girl. We have to find P (E|F).

Now $$P(F) = \frac{430}{1000} = 0.43 \text{ and } P(E \quad F) = \frac{43}{1000} \quad 0.043 \quad \text{(Why?)}$$

Then $$P(E|F) = \frac{P(E \cap F)}{P(F)} = \frac{0.043}{0.43} = 0.1$$

Example 5 A die is thrown three times. Events A and B are defined as below:

A : 4 on the third throw

B : 6 on the first and 5 on the second throw

Find the probability of A given that B has already occurred.

Solution The sample space has 216 outcomes.

Now $$A = \begin{Bmatrix} (1,1,4) & (1,2,4) & \ldots (1,6,4) & (2,1,4) & (2,2,4) & \ldots & (2,6,4) \\ (3,1,4) & (3,2,4) & \ldots (3,6,4) & (4,1,4) & (4,2,4) & \ldots & (4,6,4) \\ (5,1,4) & (5,2,4) & \ldots (5,6,4) & (6,1,4) & (6,2,4) & \ldots & (6,6,4) \end{Bmatrix}$$

$$B = \{(6,5,1), (6,5,2), (6,5,3), (6,5,4), (6,5,5), (6,5,6)\}$$

and $$A \cap B = \{(6,5,4)\}.$$

Now $$P(B) = \frac{6}{216} \text{ and } P(A \cap B) = \frac{1}{216}$$

Then $$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{1}{216}}{\frac{6}{216}} = \frac{1}{6}$$

Example 6 A die is thrown twice and the sum of the numbers appearing is observed to be 6. What is the conditional probability that the number 4 has appeared at least once?

Solution Let E be the event that 'number 4 appears at least once' and F be the event that 'the sum of the numbers appearing is 6'.

Then, $$E = \{(4,1), (4,2), (4,3), (4,4), (4,5), (4,6), (1,4), (2,4), (3,4), (5,4), (6,4)\}$$

and $$F = \{(1,5), (2,4), (3,3), (4,2), (5,1)\}$$

We have $$P(E) = \frac{11}{36} \text{ and } P(F) = \frac{5}{36}$$

Also $$E \cap F = \{(2,4), (4,2)\}$$

Therefore $\quad P(E \cap F) = \dfrac{2}{36}$

Hence, the required probability

$$P(E|F) = \frac{P(E \cap F)}{P(F)} = \frac{\frac{2}{36}}{\frac{5}{36}} = \frac{2}{5}$$

For the conditional probability discussed above, we have considered the elementary events of the experiment to be equally likely and the corresponding definition of the probability of an event was used. However, the same definition can also be used in the general case where the elementary events of the sample space are not equally likely, the probabilities $P(E \cap F)$ and $P(F)$ being calculated accordingly. Let us take up the following example.

Example 7 Consider the experiment of tossing a coin. If the coin shows head, toss it again but if it shows tail, then throw a die. Find the conditional probability of the event that 'the die shows a number greater than 4' given that 'there is at least one tail'.

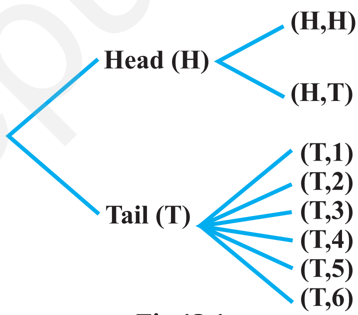


Fig 13.1

Solution The outcomes of the experiment can be represented in following diagrammatic manner called the 'tree diagram'.

The sample space of the experiment may be described as

$$S = \{(H,H), (H,T), (T,1), (T,2), (T,3), (T,4), (T,5), (T,6)\}$$

where (H, H) denotes that both the tosses result into head and (T, i) denote the first toss result into a tail and the number i appeared on the die for $i = 1,2,3,4,5,6$.

Thus, the probabilities assigned to the 8 elementary events

(H, H), (H, T), (T, 1), (T, 2), (T, 3) (T, 4), (T, 5), (T, 6)

are $\frac{1}{4}, \frac{1}{4}, \frac{1}{12}, \frac{1}{12}, \frac{1}{12}, \frac{1}{12}, \frac{1}{12}, \frac{1}{12}$ respectively which is clear from the Fig 13.2.

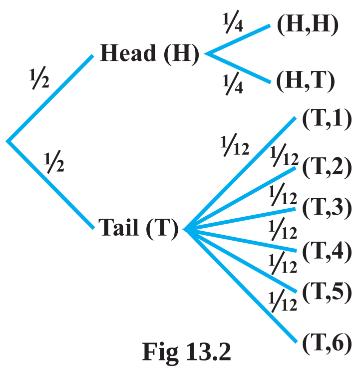


Fig 13.2

Let F be the event that 'there is at least one tail' and E be the event 'the die shows a number greater than 4'. Then

$$F = \{(H,T), (T,1), (T,2), (T,3), (T,4), (T,5), (T,6)\}$$

$$E = \{(T,5), (T,6)\} \text{ and } E \cap F = \{(T,5), (T,6)\}$$

Now $$P(F) = P(\{(H,T)\}) + P(\{(T,1)\}) + P(\{(T,2)\}) + P(\{(T,3)\}) + P(\{(T,4)\}) + P(\{(T,5)\}) + P(\{(T,6)\})$$

$$= \frac{1}{4} \quad \frac{1}{12} \quad \frac{1}{12} \quad \frac{1}{12} \quad \frac{1}{12} \quad \frac{1}{12} \quad \frac{1}{12} \quad \frac{3}{4}$$

and $$P(E \cap F) = P(\{(T,5)\}) + P(\{(T,6)\}) = \frac{1}{12} \quad \frac{1}{12} \quad \frac{1}{6}$$

Hence $$P(E|F) = \frac{P(E \cap F)}{P(F)} = \frac{\frac{1}{6}}{\frac{3}{4}} = \frac{2}{9}$$

EXERCISE 13.1

1. Given that E and F are events such that P(E) = 0.6, P(F) = 0.3 and $P(E \cap F) = 0.2$, find P(E|F) and P(F|E)
2. Compute P(A|B), if P(B) = 0.5 and $P(A \cap B) = 0.32$
3. If P(A) = 0.8, P(B) = 0.5 and P(B|A) = 0.4, find
   (i) $P(A \cap B)$ (ii) P(A|B) (iii) $P(A \cup B)$
4. Evaluate $P(A \cup B)$, if $2P(A) = P(B) = \frac{5}{13}$ and $P(A|B) = \frac{2}{5}$
5. If $P(A) = \frac{6}{11}$, $P(B) = \frac{5}{11}$ and $P(A \cup B)$ $\frac{7}{11}$, find
   (i) $P(A \cap B)$ (ii) P(A|B) (iii) P(B|A)

Determine P(E|F) in Exercises 6 to 9.

6. A coin is tossed three times, where
   (i) E : head on third toss , F : heads on first two tosses
   (ii) E : at least two heads , F : at most two heads
   (iii) E : at most two tails , F : at least one tail

7. Two coins are tossed once, where

(i) E : tail appears on one coin, F : one coin shows head

(ii) E : no tail appears, F : no head appears

8. A die is thrown three times,

E : 4 appears on the third toss, F : 6 and 5 appears respectively on first two tosses

9. Mother, father and son line up at random for a family picture

E : son on one end, F : father in middle

10. A black and a red dice are rolled.

(a) Find the conditional probability of obtaining a sum greater than 9, given that the black die resulted in a 5.

(b) Find the conditional probability of obtaining the sum 8, given that the red die resulted in a number less than 4.

11. A fair die is rolled. Consider events E = {1,3,5}, F = {2,3} and G = {2,3,4,5} Find

(i) P(E|F) and P(F|E) (ii) P(E|G) and P(G|E)

(iii) $P((E \cup F)|G)$ and $P((E \cap F)|G)$

12. Assume that each born child is equally likely to be a boy or a girl. If a family has two children, what is the conditional probability that both are girls given that (i) the youngest is a girl, (ii) at least one is a girl?

13. An instructor has a question bank consisting of 300 easy True / False questions, 200 difficult True / False questions, 500 easy multiple choice questions and 400 difficult multiple choice questions. If a question is selected at random from the question bank, what is the probability that it will be an easy question given that it is a multiple choice question?

14. Given that the two numbers appearing on throwing two dice are different. Find the probability of the event 'the sum of numbers on the dice is 4'.

15. Consider the experiment of throwing a die, if a multiple of 3 comes up, throw the die again and if any other number comes, toss a coin. Find the conditional probability of the event 'the coin shows a tail', given that 'at least one die shows a 3'.

In each of the Exercises 16 and 17 choose the correct answer:

16. If $P(A) = \frac{1}{2}$, $P(B) = 0$, then $P(A|B)$ is

(A) 0 (B) $\frac{1}{2}$

(C) not defined (D) 1

17. If A and B are events such that $P(A|B) = P(B|A)$, then

(A) $A \subset B$ but $A \neq B$ (B) $A = B$

(C) $A \cap B = \phi$ (D) $P(A) = P(B)$

13.3 Multiplication Theorem on Probability

Let E and F be two events associated with a sample space S. Clearly, the set $E \cap F$ denotes the event that both E and F have occurred. In other words, $E \cap F$ denotes the simultaneous occurrence of the events E and F. The event $E \cap F$ is also written as EF.

Very often we need to find the probability of the event EF. For example, in the experiment of drawing two cards one after the other, we may be interested in finding the probability of the event 'a king and a queen'. The probability of event EF is obtained by using the conditional probability as obtained below :

We know that the conditional probability of event E given that F has occurred is denoted by $P(E|F)$ and is given by

$$P(E|F) = \frac{P(E \cap F)}{P(F)}, P(F) \neq 0$$

From this result, we can write

$$P(E \cap F) = P(F) \,.\, P(E|F) \qquad \text{... (1)}$$

Also, we know that

$$P(F|E) = \frac{P(F \cap E)}{P(E)}, P(E) \neq 0$$

or
$$P(F|E) = \frac{P(E \cap F)}{P(E)} \quad (\text{since } E \cap F = F \cap E)$$

Thus,
$$P(E \cap F) = P(E).\ P(F|E) \qquad \text{.... (2)}$$

Combining (1) and (2), we find that

$$P(E \cap F) = P(E)\ P(F|E)$$
$$= P(F)\ P(E|F) \text{ provided } P(E) \neq 0 \text{ and } P(F) \neq 0.$$

The above result is known as the *multiplication rule of probability*.

Let us now take up an example.

Example 8 An urn contains 10 black and 5 white balls. Two balls are drawn from the urn one after the other without replacement. What is the probability that both drawn balls are black?

Solution Let E and F denote respectively the events that first and second ball drawn are black. We have to find $P(E \cap F)$ or $P(EF)$.

Now $\quad P(E) = P\text{ (black ball in first draw)} = \frac{10}{15}$

Also given that the first ball drawn is black, i.e., event E has occurred, now there are 9 black balls and five white balls left in the urn. Therefore, the probability that the second ball drawn is black, given that the ball in the first draw is black, is nothing but the conditional probability of F given that E has occurred.

i.e. $$P(F|E) = \frac{9}{14}$$

By multiplication rule of probability, we have

$$P(E \cap F) = P(E)\ P(F|E)$$

$$= \frac{10}{15}\quad \frac{9}{14}\quad \frac{3}{7}$$

Multiplication rule of probability for more than two events If E, F and G are three events of sample space, we have

$$P(E \cap F \cap G) = P(E)\ P(F|E)\ P(G|(E \cap F)) = P(E)\ P(F|E)\ P(G|EF)$$

Similarly, the multiplication rule of probability can be extended for four or more events.

The following example illustrates the extension of multiplication rule of probability for three events.

Example 9 Three cards are drawn successively, without replacement from a pack of 52 well shuffled cards. What is the probability that first two cards are kings and the third card drawn is an ace?

Solution Let K denote the event that the card drawn is king and A be the event that the card drawn is an ace. Clearly, we have to find P (KKA)

Now $$P(K) = \frac{4}{52}$$

Also, P (K|K) is the probability of second king with the condition that one king has already been drawn. Now there are three kings in (52 − 1) = 51 cards.

Therefore $$P(K|K) = \frac{3}{51}$$

Lastly, P(A|KK) is the probability of third drawn card to be an ace, with the condition that two kings have already been drawn. Now there are four aces in left 50 cards.

Therefore $$P(A|KK) = \frac{4}{50}$$

By multiplication law of probability, we have

$$P(KKA) = P(K) \quad P(K|K) \quad P(A|KK)$$

$$= \frac{4}{52} \quad \frac{3}{51} \quad \frac{4}{50} \quad \frac{2}{5525}$$

13.4 Independent Events

Consider the experiment of drawing a card from a deck of 52 playing cards, in which the elementary events are assumed to be equally likely. If E and F denote the events 'the card drawn is a spade' and 'the card drawn is an ace' respectively, then

$$P(E) = \frac{13}{52} \quad \frac{1}{4} \text{ and } P(F) \quad \frac{4}{52} \quad \frac{1}{13}$$

Also E and F is the event ' the card drawn is the ace of spades' so that

$$P(E \cap F) = \frac{1}{52}$$

Hence $$P(E|F) = \frac{P(E \quad F)}{P(F)} \quad \frac{\frac{1}{52}}{\frac{1}{13}} \quad \frac{1}{4}$$

Since $P(E) = \frac{1}{4} = P(E|F)$, we can say that the occurrence of event F has not affected the probability of occurrence of the event E.

We also have

$$P(F|E) = \frac{P(E \quad F)}{P(E)} \quad \frac{\frac{1}{52}}{\frac{1}{4}} \quad \frac{1}{13} \quad P(F)$$

Again, $P(F) = \frac{1}{13} = P(F|E)$ shows that occurrence of event E has not affected the probability of occurrence of the event F.

Thus, E and F are two events such that the probability of occurrence of one of them is not affected by occurrence of the other.

Such events are called *independent events*.

Definition 2 Two events E and F are said to be independent, if

$$P(F|E) = P(F) \text{ provided } P(E) \neq 0$$

and
$$P(E|F) = P(E) \text{ provided } P(F) \neq 0$$

Thus, in this definition we need to have $P(E) \neq 0$ and $P(F) \neq 0$

Now, by the multiplication rule of probability, we have

$$P(E \cap F) = P(E) \cdot P(F|E) \qquad \ldots (1)$$

If E and F are independent, then (1) becomes

$$P(E \cap F) = P(E) \cdot P(F) \qquad \ldots (2)$$

Thus, using (2), the independence of two events is also defined as follows:

Definition 3 Let E and F be two events associated with the same random experiment, then E and F are said to be independent if

$$P(E \cap F) = P(E) \cdot P(F)$$

Remarks

(i) Two events E and F are said to be dependent if they are not independent, i.e. if

$$P(E \cap F) \neq P(E) \cdot P(F)$$

(ii) Sometimes there is a confusion between independent events and mutually exclusive events. Term 'independent' is defined in terms of 'probability of events' whereas mutually exclusive is defined in term of events (subset of sample space). Moreover, mutually exclusive events never have an outcome common, but independent events, may have common outcome. Clearly, 'independent' and 'mutually exclusive' do not have the same meaning.

In other words, two independent events having nonzero probabilities of occurrence can not be mutually exclusive, and conversely, i.e. two mutually exclusive events having nonzero probabilities of occurrence can not be independent.

(iii) Two experiments are said to be independent if for every pair of events E and F, where E is associated with the first experiment and F with the second experiment, the probability of the simultaneous occurrence of the events E and F when the two experiments are performed is the product of P(E) and P(F) calculated separately on the basis of two experiments, i.e., $P(E \cap F) = P(E) \cdot P(F)$

(iv) Three events A, B and C are said to be mutually independent, if

$$P(A \cap B) = P(A)\, P(B)$$

$$P(A \cap C) = P(A)\, P(C)$$

$$P(B \cap C) = P(B)\, P(C)$$

and
$$P(A \cap B \cap C) = P(A)\, P(B)\, P(C)$$

If at least one of the above is not true for three given events, we say that the events are not independent.

Example 10 A die is thrown. If E is the event 'the number appearing is a multiple of 3' and F be the event 'the number appearing is even' then find whether E and F are independent ?

Solution We know that the sample space is S = {1, 2, 3, 4, 5, 6}

Now $\quad E = \{3, 6\}, F = \{2, 4, 6\}$ and $E \cap F = \{6\}$

Then $\quad P(E) = \frac{2}{6} = \frac{1}{3}, \; P(F) = \frac{3}{6} = \frac{1}{2}$ and $P(E \cap F) = \frac{1}{6}$

Clearly $\quad P(E \cap F) = P(E).\, P(F)$

Hence $\quad$ E and F are independent events.

Example 11 An unbiased die is thrown twice. Let the event A be 'odd number on the first throw' and B the event 'odd number on the second throw'. Check the independence of the events A and B.

Solution If all the 36 elementary events of the experiment are considered to be equally likely, we have

$$P(A) = \frac{18}{36} = \frac{1}{2} \text{ and } P(B) \; \frac{18}{36} \; \frac{1}{2}$$

Also $\quad P(A \cap B) = P$ (odd number on both throws)

$$= \frac{9}{36} = \frac{1}{4}$$

Now $\quad P(A)\, P(B) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$

Clearly $\quad P(A \cap B) = P(A) \times P(B)$

Thus, $\quad$ A and B are independent events

Example 12 Three coins are tossed simultaneously. Consider the event E 'three heads or three tails', F 'at least two heads' and G 'at most two heads'. Of the pairs (E,F), (E,G) and (F,G), which are independent? which are dependent?

Solution The sample space of the experiment is given by

$$S = \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}$$

Clearly $\quad E = \{HHH, TTT\}, F = \{HHH, HHT, HTH, THH\}$

and $\qquad G = \{HHT, HTH, THH, HTT, THT, TTH, TTT\}$

Also $\qquad E \cap F = \{HHH\}, E \cap G = \{TTT\}, F \cap G = \{ HHT, HTH, THH\}$

Therefore $\qquad P(E) = \frac{2}{8} = \frac{1}{4}, P(F) = \frac{4}{8} = \frac{1}{2}, P(G) = \frac{7}{8}$

and $\qquad P(E \cap F) = \frac{1}{8}, P(E \cap G) = \frac{1}{8}, P(F \cap G) = \frac{3}{8}$

Also $\qquad P(E) \,.\, P(F) = \frac{1}{4} \;\; \frac{1}{2} \;\; \frac{1}{8}, P(E) \;\; P(G) \;\; \frac{1}{4} \;\; \frac{7}{8} \;\; \frac{7}{32}$

and $\qquad P(F) \,.\, P(G) = \frac{1}{2} \;\; \frac{7}{8} \;\; \frac{7}{16}$

Thus $\qquad P(E \cap F) = P(E) \,.\, P(F)$

$\qquad P(E \cap G) \neq P(E) \,.\, P(G)$

and $\qquad P(F \cap G) \neq P(F) \,.\, P(G)$

Hence, the events (E and F) are independent, and the events (E and G) and (F and G) are dependent.

Example 13 Prove that if E and F are independent events, then so are the events E and F′.

Solution Since E and F are independent, we have

$$P(E \cap F) = P(E) \,.\, P(F) \qquad \text{....(1)}$$

From the venn diagram in Fig 13.3, it is clear that $E \cap F$ and $E \cap F'$ are mutually exclusive events and also $E = (E \cap F) \cup (E \cap F')$.

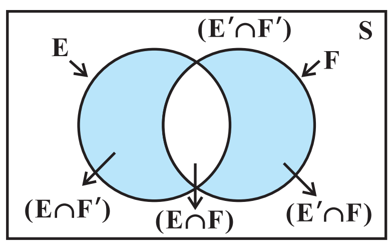


Fig 13.3

Therefore $\qquad P(E) = P(E \cap F) + P(E \cap F')$

or $\qquad P(E \cap F') = P(E) - P(E \cap F)$

$\qquad = P(E) - P(E) \,.\, P(F)$ (by (1))

$\qquad = P(E)\ (1 - P(F))$

$\qquad = P(E).\ P(F')$

Hence, E and F′ are independent

Note In a similar manner, it can be shown that if the events E and F are independent, then

(a) E′ and F are independent,

(b) E′ and F′ are independent

Example 14 If A and B are two independent events, then the probability of occurrence of at least one of A and B is given by 1– P(A′) P(B′)

Solution We have

$$
\begin{aligned}
\text{P(at least one of A and B)} &= P(A \cup B)\\
&= P(A) + P(B) - P(A \cap B)\\
&= P(A) + P(B) - P(A)\, P(B)\\
&= P(A) + P(B)\, [1-P(A)]\\
&= P(A) + P(B).\, P(A')\\
&= 1 - P(A') + P(B)\, P(A')\\
&= 1 - P(A')\, [1 - P(B)]\\
&= 1 - P(A')\, P(B')
\end{aligned}
$$

EXERCISE 13.2

1. If P(A) $\frac{3}{5}$ and P (B) $\frac{1}{5}$, find P (A $\cap$ B) if A and B are independent events.

2. Two cards are drawn at random and without replacement from a pack of 52 playing cards. Find the probability that both the cards are black.

3. A box of oranges is inspected by examining three randomly selected oranges drawn without replacement. If all the three oranges are good, the box is approved for sale, otherwise, it is rejected. Find the probability that a box containing 15 oranges out of which 12 are good and 3 are bad ones will be approved for sale.

4. A fair coin and an unbiased die are tossed. Let A be the event 'head appears on the coin' and B be the event '3 on the die'. Check whether A and B are independent events or not.

5. A die marked 1, 2, 3 in red and 4, 5, 6 in green is tossed. Let A be the event, 'the number is even,' and B be the event, 'the number is red'. Are A and B independent?

6. Let E and F be events with P(E) $\frac{3}{5}$, P(F) $= \frac{3}{10}$ and P (E $\cap$ F) = $\frac{1}{5}$. Are E and F independent?

7. Given that the events A and B are such that $P(A) = \frac{1}{2}$, $P(A \cup B) = \frac{3}{5}$ and $P(B) = p$. Find p if they are (i) mutually exclusive (ii) independent.

8. Let A and B be independent events with P(A) = 0.3 and P(B) = 0.4. Find
   (i) $P(A \cap B)$ (ii) $P(A \cup B)$
   (iii) $P(A|B)$ (iv) $P(B|A)$

9. If A and B are two events such that $P(A) = \frac{1}{4}$, $P(B) = \frac{1}{2}$ and $P(A \cap B) = \frac{1}{8}$, find P (not A and not B).

10. Events A and B are such that $P(A) = \frac{1}{2}$, $P(B) = \frac{7}{12}$ and P(not A or not B) $= \frac{1}{4}$. State whether A and B are independent ?

11. Given two independent events A and B such that P(A) = 0.3, P(B) = 0.6. Find
    (i) P(A and B) (ii) P(A and not B)
    (iii) P(A or B) (iv) P(neither A nor B)

12. A die is tossed thrice. Find the probability of getting an odd number at least once.

13. Two balls are drawn at random with replacement from a box containing 10 black and 8 red balls. Find the probability that
    (i) both balls are red.
    (ii) first ball is black and second is red.
    (iii) one of them is black and other is red.

14. Probability of solving specific problem independently by A and B are $\frac{1}{2}$ and $\frac{1}{3}$ respectively. If both try to solve the problem independently, find the probability that
    (i) the problem is solved (ii) exactly one of them solves the problem.

15. One card is drawn at random from a well shuffled deck of 52 cards. In which of the following cases are the events E and F independent ?
    (i) E : 'the card drawn is a spade'
    F : 'the card drawn is an ace'
    (ii) E : 'the card drawn is black'
    F : 'the card drawn is a king'
    (iii) E : 'the card drawn is a king or queen'
    F : 'the card drawn is a queen or jack'.

16. In a hostel, 60% of the students read Hindi news paper, 40% read English news paper and 20% read both Hindi and English news papers. A student is selected at random.

(a) Find the probability that she reads neither Hindi nor English news papers.

(b) If she reads Hindi news paper, find the probability that she reads English news paper.

(c) If she reads English news paper, find the probability that she reads Hindi news paper.

Choose the correct answer in Exercises 17 and 18.

17. The probability of obtaining an even prime number on each die, when a pair of dice is rolled is

(A) 0 (B) $\frac{1}{3}$ (C) $\frac{1}{12}$ (D) $\frac{1}{36}$

18. Two events A and B will be independent, if

(A) A and B are mutually exclusive

(B) $P(A'B') = [1 - P(A)]\ [1 - P(B)]$

(C) $P(A) = P(B)$

(D) $P(A) + P(B) = 1$

13.5 Bayes' Theorem

Consider that there are two bags I and II. Bag I contains 2 white and 3 red balls and Bag II contains 4 white and 5 red balls. One ball is drawn at random from one of the bags. We can find the probability of selecting any of the bags (i.e. $\frac{1}{2}$) or probability of drawing a ball of a particular colour (say white) from a particular bag (say Bag I). In other words, we can find the probability that the ball drawn is of a particular colour, if we are given the bag from which the ball is drawn. But, can we find the probability that the ball drawn is from a particular bag (say Bag II), if the colour of the ball drawn is given? Here, we have to find the reverse probability of Bag II to be selected when an event occurred after it is known. Famous mathematician, John Bayes' solved the problem of finding reverse probability by using conditional probability. The formula developed by him is known as *'Bayes theorem'* which was published posthumously in 1763. Before stating and proving the Bayes' theorem, let us first take up a definition and some preliminary results.

13.5.1 *Partition of a sample space*

A set of events $E_1, E_2, ..., E_n$ is said to represent a partition of the sample space S if

(a) $E_i \cap E_j = \phi,\ i \neq j,\ i, j = 1, 2, 3, ..., n$

(b) $E_1 \cup E_2 \cup ... \cup E_n = S$ and

(c) $P(E_i) > 0$ for all $i = 1, 2, ..., n$.

In other words, the events $E_1, E_2, ..., E_n$ represent a partition of the sample space S if they are pairwise disjoint, exhaustive and have nonzero probabilities.

As an example, we see that any nonempty event E and its complement E′ form a partition of the sample space S since they satisfy $E \cap E' = \phi$ and $E \cup E' = S$.

From the Venn diagram in Fig 13.3, one can easily observe that if E and F are any two events associated with a sample space S, then the set $\{E \cap F', E \cap F, E' \cap F, E' \cap F'\}$ is a partition of the sample space S. It may be mentioned that the partition of a sample space is not unique. There can be several partitions of the same sample space.

We shall now prove a theorem known as *Theorem of total probability*.

13.5.2 *Theorem of total probability*

Let $\{E_1, E_2,...,E_n\}$ be a partition of the sample space S, and suppose that each of the events $E_1, E_2,..., E_n$ has nonzero probability of occurrence. Let A be any event associated with S, then

$$P(A) = P(E_1)\ P(A|E_1) + P(E_2)\ P(A|E_2) + ... + P(E_n)\ P(A|E_n)$$

$$= \sum_{j=1}^{n} P(E_j) P(A|E_j)$$

Proof Given that $E_1, E_2,..., E_n$ is a partition of the sample space S (Fig 13.4). Therefore,

$$S = E_1 \cup E_2 \cup ... \cup E_n \quad ... (1)$$

and $\qquad E_i \cap E_j = \phi,\ i \neq j,\ i, j = 1, 2, ..., n$

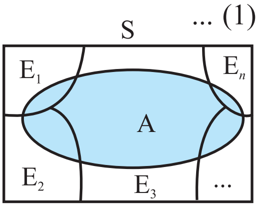


Fig 13.4

Now, we know that for any event A,

$$A = A \cap S$$
$$= A \cap (E_1 \cup E_2 \cup ... \cup E_n)$$
$$= (A \cap E_1) \cup (A \cap E_2) \cup ... \cup (A \cap E_n)$$

Also $A \cap E_i$ and $A \cap E_j$ are respectively the subsets of E_i and E_j. We know that E_i and E_j are disjoint, for $i \neq j$, therefore, $A \cap E_i$ and $A \cap E_j$ are also disjoint for all $i \neq j,\ i, j = 1, 2, ..., n$.

Thus, $\qquad P(A) = P\ [(A \cap E_1) \cup (A \cap E_2) \cup ..... \cup (A \cap E_n)]$

$$= P\ (A \cap E_1) + P\ (A \cap E_2) + ... + P\ (A \cap E_n)$$

Now, by multiplication rule of probability, we have

$$P(A \cap E_i) = P(E_i)\ P(A|E_i) \text{ as } P\ (E_i) \neq 0 \ \forall i = 1,2,..., n$$

Therefore, $\quad P(A) = P(E_1)\, P(A|E_1) + P(E_2)\, P(A|E_2) + ... + P(E_n)P(A|E_n)$

or $\quad P(A) = \sum_{j=1}^{n} P(E_j)P(A|E_j)$

Example 15 A person has undertaken a construction job. The probabilities are 0.65 that there will be strike, 0.80 that the construction job will be completed on time if there is no strike, and 0.32 that the construction job will be completed on time if there is a strike. Determine the probability that the construction job will be completed on time.

Solution Let A be the event that the construction job will be completed on time, and B be the event that there will be a strike. We have to find P(A).
We have

$$P(B) = 0.65,\ P(\text{no strike}) = P(B') = 1 - P(B) = 1 - 0.65 = 0.35$$

$$P(A|B) = 0.32,\ P(A|B') = 0.80$$

Since events B and B′ form a partition of the sample space S, therefore, by theorem on total probability, we have

$$P(A) = P(B)\ P(A|B) + P(B')\ P(A|B')$$
$$= 0.65 \times 0.32 + 0.35 \times 0.8$$
$$= 0.208 + 0.28 = 0.488$$

Thus, the probability that the construction job will be completed in time is 0.488.

We shall now state and prove the Bayes' theorem.

Bayes' Theorem If $E_1, E_2, ..., E_n$ are n non empty events which constitute a partition of sample space S, i.e. $E_1, E_2, ..., E_n$ are pairwise disjoint and $E_1 \cup E_2 \cup ... \cup E_n = S$ and A is any event of nonzero probability, then

$$P(E_i|A) = \frac{P(E_i)P(A|E_i)}{\sum_{j=1}^{n} P(E_j)P(A|E_j)} \quad \text{for any } i = 1, 2, 3, ..., n$$

Proof By formula of conditional probability, we know that

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

$$= \frac{P(E_i)P(A|E_i)}{P(A)} \quad \text{(by multiplication rule of probability)}$$

$$= \frac{P(E_i)P(A|E_i)}{\sum_{j=1}^{n} P(E_j)P(A|E_j)} \quad \text{(by the result of theorem of total probability)}$$

Remark The following terminology is generally used when Bayes' theorem is applied.

The events $E_1, E_2, ..., E_n$ are called *hypotheses.*

The probability $P(E_i)$ is called the *priori probability* of the hypothesis E_i

The conditional probability $P(E_i|A)$ is called *a posteriori probability* of the hypothesis E_i.

Bayes' theorem is also called the formula for the probability of "causes". Since the E_i's are a partition of the sample space S, one and only one of the events E_i occurs (i.e. one of the events E_i must occur and only one can occur). Hence, the above formula gives us the probability of a particular E_i (i.e. a "Cause"), given that the event A has occurred.

The Bayes' theorem has its applications in variety of situations, few of which are illustrated in following examples.

Example 16 Bag I contains 3 red and 4 black balls while another Bag II contains 5 red and 6 black balls. One ball is drawn at random from one of the bags and it is found to be red. Find the probability that it was drawn from Bag II.

Solution Let E_1 be the event of choosing the bag I, E_2 the event of choosing the bag II and A be the event of drawing a red ball.

Then $$P(E_1) = P(E_2) = \frac{1}{2}$$

Also $$P(A|E_1) = P(\text{drawing a red ball from Bag I}) = \frac{3}{7}$$

and $$P(A|E_2) = P(\text{drawing a red ball from Bag II}) = \frac{5}{11}$$

Now, the probability of drawing a ball from Bag II, being given that it is red, is $P(E_2|A)$

By using Bayes' theorem, we have

$$P(E_2|A) = \frac{P(E_2)P(A|E_2)}{P(E_1)P(A|E_1)+P(E_2)P(A|E_2)} = \frac{\frac{1}{2}\times\frac{5}{11}}{\frac{1}{2}\times\frac{3}{7}+\frac{1}{2}\times\frac{5}{11}} = \frac{35}{68}$$

Example 17 Given three identical boxes I, II and III, each containing two coins. In box I, both coins are gold coins, in box II, both are silver coins and in the box III, there is one gold and one silver coin. A person chooses a box at random and takes out a coin. If the coin is of gold, what is the probability that the other coin in the box is also of gold?

Solution Let E_1, E_2 and E_3 be the events that boxes I, II and III are chosen, respectively.

Then $$P(E_1) = P(E_2) = P(E_3) = \frac{1}{3}$$

Also, let A be the event that 'the coin drawn is of gold'

Then $$P(A|E_1) = P(\text{a gold coin from bag I}) = \frac{2}{2} = 1$$

$$P(A|E_2) = P(\text{a gold coin from bag II}) = 0$$

$$P(A|E_3) = P(\text{a gold coin from bag III}) = \frac{1}{2}$$

Now, the probability that the other coin in the box is of gold

= the probability that gold coin is drawn from the box I.

$= P(E_1|A)$

By Bayes' theorem, we know that

$$P(E_1|A) = \frac{P(E_1)P(A|E_1)}{P(E_1)P(A|E_1)+P(E_2)P(A|E_2)+P(E_3)P(A|E_3)}$$

$$= \frac{\frac{1}{3}\times 1}{\frac{1}{3}\times 1+\frac{1}{3}\times 0+\frac{1}{3}\times\frac{1}{2}} = \frac{2}{3}$$

Example 18 Suppose that the reliability of a HIV test is specified as follows:

Of people having HIV, 90% of the test detect the disease but 10% go undetected. Of people free of HIV, 99% of the test are judged HIV–ive but 1% are diagnosed as showing HIV+ive. From a large population of which only 0.1% have HIV, one person is selected at random, given the HIV test, and the pathologist reports him/her as HIV+ive. What is the probability that the person actually has HIV?

Solution Let E denote the event that the person selected is actually having HIV and A the event that the person's HIV test is diagnosed as +ive. We need to find P(E|A).

Also E′ denotes the event that the person selected is actually not having HIV.

Clearly, {E, E′} is a partition of the sample space of all people in the population. We are given that

$$P(E) = 0.1\% \quad \frac{0.1}{100} \quad 0.001$$

$$P(E') = 1 - P(E) = 0.999$$

$P(A|E)$ = P(Person tested as HIV+ive given that he/she is actually having HIV)

$$= 90\% \quad \frac{90}{100} \quad 0.9$$

and $P(A|E')$ = P(Person tested as HIV +ive given that he/she is actually not having HIV)

$$= 1\% = \frac{1}{100} = 0.01$$

Now, by Bayes' theorem

$$P(E|A) = \frac{P(E)P(A|E)}{P(E)P(A|E)+P(E\ )P(A|E\ )}$$

$$= \frac{0.001\times0.9}{0.001\times0.9+0.999\times0.01} = \frac{90}{1089}$$

$$= 0.083 \text{ approx.}$$

Thus, the probability that a person selected at random is actually having HIV given that he/she is tested HIV+ive is 0.083.

Example 19 In a factory which manufactures bolts, machines A, B and C manufacture respectively 25%, 35% and 40% of the bolts. Of their outputs, 5, 4 and 2 percent are respectively defective bolts. A bolt is drawn at random from the product and is found to be defective. What is the probability that it is manufactured by the machine B?

Solution Let events B_1, B_2, B_3 be the following :

B_1 : the bolt is manufactured by machine A

B_2 : the bolt is manufactured by machine B

B_3 : the bolt is manufactured by machine C

Clearly, B_1, B_2, B_3 are mutually exclusive and exhaustive events and hence, they represent a partition of the sample space.

Let the event E be 'the bolt is defective'.

The event E occurs with B_1 or with B_2 or with B_3. Given that,

$$P(B_1) = 25\% = 0.25,\ \ P(B_2) = 0.35 \text{ and } P(B_3) = 0.40$$

Again $P(E|B_1)$ = Probability that the bolt drawn is defective given that it is manufactured by machine A = 5% = 0.05

Similarly, $P(E|B_2) = 0.04$, $P(E|B_3) = 0.02$.

Hence, by Bayes' Theorem, we have

$$P(B_2|E) = \frac{P(B_2)P(E|B_2)}{P(B_1)P(E|B_1)+P(B_2)P(E|B_2)+P(B_3)P(E|B_3)}$$

$$= \frac{0.35\times 0.04}{0.25\times 0.05+0.35\times 0.04+0.40\times 0.02}$$

$$= \frac{0.0140}{0.0345}=\frac{28}{69}$$

Example 20 A doctor is to visit a patient. From the past experience, it is known that the probabilities that he will come by train, bus, scooter or by other means of transport are respectively $\frac{3}{10}, \frac{1}{5}, \frac{1}{10}$ and $\frac{2}{5}$. The probabilities that he will be late are $\frac{1}{4}, \frac{1}{3}$, and $\frac{1}{12}$, if he comes by train, bus and scooter respectively, but if he comes by other means of transport, then he will not be late. When he arrives, he is late. What is the probability that he comes by train?

Solution Let E be the event that the doctor visits the patient late and let T_1, T_2, T_3, T_4 be the events that the doctor comes by train, bus, scooter, and other means of transport respectively.

Then $$P(T_1) = \frac{3}{10}, P(T_2)=\frac{1}{5}, P(T_3)=\frac{1}{10} \text{ and } P(T_4)=\frac{2}{5} \quad \text{(given)}$$

$P(E|T_1)$ = Probability that the doctor arriving late comes by train = $\frac{1}{4}$

Similarly, $P(E|T_2) = \frac{1}{3}$, $P(E|T_3) = \frac{1}{12}$ and $P(E|T_4) = 0$, since he is not late if he comes by other means of transport.

Therefore, by Bayes' Theorem, we have

$P(T_1|E)$ = Probability that the doctor arriving late comes by train

$$= \frac{P(T_1)P(E|T_1)}{P(T_1)P(E|T_1)+P(T_2)P(E|T_2)+P(T_3)P(E|T_3)+P(T_4)P(E|T_4)}$$

$$= \frac{\frac{3}{10}\ \frac{1}{4}}{\frac{3}{10}\ \frac{1}{4}\ \frac{1}{5}\ \frac{1}{3}\ \frac{1}{10}\ \frac{1}{12}\ \frac{2}{5}\ 0} = \frac{3}{40}\times\frac{120}{18}=\frac{1}{2}$$

Hence, the required probability is $\frac{1}{2}$.

Example 21 A man is known to speak truth 3 out of 4 times. He throws a die and reports that it is a six. Find the probability that it is actually a six.

Solution Let E be the event that the man reports that six occurs in the throwing of the die and let S_1 be the event that six occurs and S_2 be the event that six does not occur.

Then $\qquad P(S_1)$ = Probability that *six* occurs = $\frac{1}{6}$

$P(S_2)$ = Probability that *six* does not occur = $\frac{5}{6}$

$P(E|S_1)$ = Probability that the man reports that *six* occurs when *six* has actually occurred on the die

= Probability that the man speaks the truth = $\frac{3}{4}$

$P(E|S_2)$ = Probability that the man reports that *six* occurs when *six* has not actually occurred on the die

= Probability that the man does not speak the truth $1 \quad \frac{3}{4} \quad \frac{1}{4}$

Thus, by Bayes' theorem, we get

$P(S_1|E)$ = Probability that the report of the man that *six* has occurred is actually a *six*

$$= \frac{P(S_1)P(E|S_1)}{P(S_1)P(E|S_1)+P(S_2)P(E|S_2)}$$

$$= \frac{\frac{1}{6} \quad \frac{3}{4}}{\frac{1}{6} \quad \frac{3}{4} \quad \frac{5}{6} \quad \frac{1}{4}} \quad \frac{1}{8} \quad \frac{24}{8} \quad \frac{3}{8}$$

Hence, the required probability is $\frac{3}{8}$.

EXERCISE 13.3

1. An urn contains 5 red and 5 black balls. A ball is drawn at random, its colour is noted and is returned to the urn. Moreover, 2 additional balls of the colour drawn are put in the urn and then a ball is drawn at random. What is the probability that the second ball is red?

2. A bag contains 4 red and 4 black balls, another bag contains 2 red and 6 black balls. One of the two bags is selected at random and a ball is drawn from the bag which is found to be red. Find the probability that the ball is drawn from the first bag.
3. Of the students in a college, it is known that 60% reside in hostel and 40% are day scholars (not residing in hostel). Previous year results report that 30% of all students who reside in hostel attain A grade and 20% of day scholars attain A grade in their annual examination. At the end of the year, one student is chosen at random from the college and he has an A grade, what is the probability that the student is a hostlier?
4. In answering a question on a multiple choice test, a student either knows the answer or guesses. Let $\frac{3}{4}$ be the probability that he knows the answer and $\frac{1}{4}$ be the probability that he guesses. Assuming that a student who guesses at the answer will be correct with probability $\frac{1}{4}$. What is the probability that the student knows the answer given that he answered it correctly?
5. A laboratory blood test is 99% effective in detecting a certain disease when it is in fact, present. However, the test also yields a false positive result for 0.5% of the healthy person tested (i.e. if a healthy person is tested, then, with probability 0.005, the test will imply he has the disease). If 0.1 percent of the population actually has the disease, what is the probability that a person has the disease given that his test result is positive ?
6. There are three coins. One is a two headed coin (having head on both faces), another is a biased coin that comes up heads 75% of the time and third is an unbiased coin. One of the three coins is chosen at random and tossed, it shows heads, what is the probability that it was the two headed coin ?
7. An insurance company insured 2000 scooter drivers, 4000 car drivers and 6000 truck drivers. The probability of an accidents are 0.01, 0.03 and 0.15 respectively. One of the insured persons meets with an accident. What is the probability that he is a scooter driver?
8. A factory has two machines A and B. Past record shows that machine A produced 60% of the items of output and machine B produced 40% of the items. Further, 2% of the items produced by machine A and 1% produced by machine B were defective. All the items are put into one stockpile and then one item is chosen at random from this and is found to be defective. What is the probability that it was produced by machine B?
9. Two groups are competing for the position on the Board of directors of a corporation. The probabilities that the first and the second groups will win are

0.6 and 0.4 respectively. Further, if the first group wins, the probability of introducing a new product is 0.7 and the corresponding probability is 0.3 if the second group wins. Find the probability that the new product introduced was by the second group.

10. Suppose a girl throws a die. If she gets a 5 or 6, she tosses a coin three times and notes the number of heads. If she gets 1, 2, 3 or 4, she tosses a coin once and notes whether a head or tail is obtained. If she obtained exactly one head, what is the probability that she threw 1, 2, 3 or 4 with the die?

11. A manufacturer has three machine operators A, B and C. The first operator A produces 1% defective items, where as the other two operators B and C produce 5% and 7% defective items respectively. A is on the job for 50% of the time, B is on the job for 30% of the time and C is on the job for 20% of the time. A defective item is produced, what is the probability that it was produced by A?

12. A card from a pack of 52 cards is lost. From the remaining cards of the pack, two cards are drawn and are found to be both diamonds. Find the probability of the lost card being a diamond.

13. Probability that A speaks truth is $\frac{4}{5}$. A coin is tossed. A reports that a head appears. The probability that actually there was head is

(A) $\frac{4}{5}$ (B) $\frac{1}{2}$ (C) $\frac{1}{5}$ (D) $\frac{2}{5}$

14. If A and B are two events such that $A \subset B$ and $P(B) \neq 0$, then which of the following is correct?

(A) $P(A|B)\ \frac{P(B)}{P(A)}$ (B) $P(A|B) < P(A)$

(C) $P(A|B) \geq P(A)$ (D) None of these

13.6 Random Variables and its Probability Distributions

We have already learnt about random experiments and formation of sample spaces. In most of these experiments, we were not only interested in the particular outcome that occurs but rather in some number associated with that outcomes as shown in following examples/experiments.

(i) In tossing two dice, we may be interested in the sum of the numbers on the two dice.

(ii) In tossing a coin 50 times, we may want the number of heads obtained.

(iii) In the experiment of taking out four articles (one after the other) at random from a lot of 20 articles in which 6 are defective, we want to know the number of defectives in the sample of four and not in the particular sequence of defective and nondefective articles.

In all of the above experiments, we have a rule which assigns to each outcome of the experiment a single real number. This single real number may vary with different outcomes of the experiment. Hence, it is a variable. Also its value depends upon the outcome of a random experiment and, hence, is called random variable. A random variable is usually denoted by X.

If you recall the definition of a function, you will realise that the random variable X is really speaking a function whose domain is the set of outcomes (or sample space) of a random experiment. A random variable can take any real value, therefore, its co-domain is the set of real numbers. Hence, a random variable can be defined as follows :

Definition 4 A random variable is a real valued function whose domain is the sample space of a random experiment.

For example, let us consider the experiment of tossing a coin two times in succession.

The sample space of the experiment is S = {HH, HT, TH, TT}.

If X denotes the number of heads obtained, then X is a random variable and for each outcome, its value is as given below :

$$X(HH) = 2,\ X(HT) = 1,\ X(TH) = 1,\ X(TT) = 0.$$

More than one random variables can be defined on the same sample space. For example, let Y denote the number of heads minus the number of tails for each outcome of the above sample space S.

Then $$Y(HH) = 2,\ Y(HT) = 0,\ Y(TH) = 0,\ Y(TT) = -2.$$

Thus, X and Y are two different random variables defined on the same sample space S.

Example 22 A person plays a game of tossing a coin thrice. For each head, he is given Rs 2 by the organiser of the game and for each tail, he has to give Rs 1.50 to the organiser. Let X denote the amount gained or lost by the person. Show that X is a random variable and exhibit it as a function on the sample space of the experiment.

Solution X is a number whose values are defined on the outcomes of a random experiment. Therefore, X is a random variable.

Now, sample space of the experiment is

$$S = \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}$$

Then $X(HHH) = \text{Rs } (2 \times 3) = \text{Rs } 6$

$X(HHT) = X(HTH) = X(THH) = \text{Rs } (2 \times 2 - 1 \times 1.50) = \text{Rs } 2.50$

$X(HTT) = X(THT) = (TTH) = \text{Rs } (1 \times 2) - (2 \times 1.50) = - \text{Re } 1$

and $X(TTT) = - \text{Rs } (3 \times 1.50) = - \text{Rs } 4.50$

where, minus sign shows the loss to the player. Thus, for each element of the sample space, X takes a unique value, hence, X is a function on the sample space whose range is

$$\{-1,\ 2.50,\ -4.50,\ 6\}$$

Example 23 A bag contains 2 white and 1 red balls. One ball is drawn at random and then put back in the box after noting its colour. The process is repeated again. If X denotes the number of red balls recorded in the two draws, describe X.

Solution Let the balls in the bag be denoted by w_1, w_2, r. Then the sample space is

$$S = \{w_1\, w_1,\ w_1\, w_2,\ w_2\, w_2,\ w_2\, w_1,\ w_1\, r,\ w_2\, r,\ r\, w_1,\ r\, w_2,\ r\, r\}$$

Now, for $\omega \in S$

$$X(\omega) = \text{number of red balls}$$

Therefore

$$X(\{w_1\, w_1\}) = X(\{w_1\, w_2\}) = X(\{w_2\, w_2\}) = X(\{w_2\, w_1\}) = 0$$

$$X(\{w_1\, r\}) = X(\{w_2\, r\}) = X(\{r\, w_1\}) = X(\{r\, w_2\}) = 1 \text{ and } X(\{r\, r\}) = 2$$

Thus, X is a random variable which can take values 0, 1 or 2.

13.6.1 *Probability distribution of a random variable*

Let us look at the experiment of selecting one family out of ten families $f_1, f_2, ..., f_{10}$ in such a manner that each family is equally likely to be selected. Let the families $f_1, f_2, ..., f_{10}$ have 3, 4, 3, 2, 5, 4, 3, 6, 4, 5 members, respectively.

Let us select a family and note down the number of members in the family denoting X. Clearly, X is a random variable defined as below :

$$X(f_1) = 3,\ X(f_2) = 4,\ X(f_3) = 3,\ X(f_4) = 2,\ X(f_5) = 5,$$

$$X(f_6) = 4,\ X(f_7) = 3,\ X(f_8) = 6,\ X(f_9) = 4,\ X(f_{10}) = 5$$

Thus, X can take any value 2,3,4,5 or 6 depending upon which family is selected.

Now, X will take the value 2 when the family f_4 is selected. X can take the value 3 when any one of the families f_1, f_3, f_7 is selected.

Similarly, $X = 4$, when family f_2, f_6 or f_9 is selected,

$X = 5$, when family f_5 or f_{10} is selected

and $X = 6$, when family f_8 is selected.

Since we had assumed that each family is equally likely to be selected, the probability that family f_4 is selected is $\frac{1}{10}$.

Thus, the probability that X can take the value 2 is $\frac{1}{10}$. We write $P(X = 2) = \frac{1}{10}$

Also, the probability that any one of the families f_1, f_3 or f_7 is selected is

$$P(\{f_1, f_3, f_7\}) = \frac{3}{10}$$

Thus, the probability that X can take the value 3 = $\frac{3}{10}$

We write $$P(X = 3) = \frac{3}{10}$$

Similarly, we obtain

$$P(X = 4) = P(\{f_2, f_6, f_9\}) = \frac{3}{10}$$

$$P(X = 5) = P(\{f_5, f_{10}\}) = \frac{2}{10}$$

and $$P(X = 6) = P(\{f_8\}) = \frac{1}{10}$$

Such a description giving the values of the random variable along with the corresponding probabilities is called the *probability distribution of the random variable X*.

In general, the probability distribution of a random variable X is defined as follows:

Definition 5 The probability distribution of a random variable X is the system of numbers

X	:	x_1	x_2	...	x_n
P(X)	:	p_1	p_2	...	p_n

where, $$p_i \geq 0, \quad \sum_{i=1}^{n} p_i = 1, \; i = 1, 2, ..., n$$

The real numbers $x_1, x_2, ..., x_n$ are the possible values of the random variable X and p_i (i = 1,2,..., n) is the probability of the random variable X taking the value x_i i.e., $P(X = x_i) = p_i$

> **Note** If x_i is one of the possible values of a random variable X, the statement $X = x_i$ is true only at some point (s) of the sample space. Hence, the probability that X takes value x_i is always nonzero, i.e. $P(X = x_i) \neq 0$.

Also for all possible values of the random variable X, all elements of the sample space are covered. Hence, the sum of all the probabilities in a probability distribution must be one.

Example 24 Two cards are drawn successively with replacement from a well-shuffled deck of 52 cards. Find the probability distribution of the number of aces.

Solution The number of aces is a random variable. Let it be denoted by X. Clearly, X can take the values 0, 1, or 2.

Now, since the draws are done with replacement, therefore, the two draws form independent experiments.

Therefore,
$$\begin{aligned} P(X = 0) &= P(\text{non-ace and non-ace}) \\ &= P(\text{non-ace}) \times P(\text{non-ace}) \\ &= \frac{48}{52} \times \frac{48}{52} = \frac{144}{169} \end{aligned}$$

$$\begin{aligned} P(X = 1) &= P(\text{ace and non-ace or non-ace and ace}) \\ &= P(\text{ace and non-ace}) + P(\text{non-ace and ace}) \\ &= P(\text{ace}).\ P(\text{non-ace}) + P\ (\text{non-ace})\ .\ P(\text{ace}) \\ &= \frac{4}{52} \times \frac{48}{52} + \frac{48}{52} \times \frac{4}{52} = \frac{24}{169} \end{aligned}$$

and
$$\begin{aligned} P(X = 2) &= P\ (\text{ace and ace}) \\ &= \frac{4}{52} \quad \frac{4}{52} \quad \frac{1}{169} \end{aligned}$$

Thus, the required probability distribution is

X	0	1	2
P(X)	$\frac{144}{169}$	$\frac{24}{169}$	$\frac{1}{169}$

Example 25 Find the probability distribution of number of doublets in three throws of a pair of dice.

Solution Let X denote the number of doublets. Possible doublets are

$$(1,1)\ ,(2,2),\ (3,3),\ (4,4),\ (5,5),(6,6)$$

Clearly, X can take the value 0, 1, 2, or 3.

Probability of getting a doublet $\frac{6}{36}\quad\frac{1}{6}$

Probability of not getting a doublet $1\quad\frac{1}{6}\quad\frac{5}{6}$

Now $\quad$ P(X = 0) = P (no doublet) = $\frac{5}{6}\quad\frac{5}{6}\quad\frac{5}{6}\quad\frac{125}{216}$

P(X = 1) = P (one doublet and two non-doublets)

$$= \frac{1}{6}\quad\frac{5}{6}\quad\frac{5}{6}\quad\frac{5}{6}\quad\frac{1}{6}\quad\frac{5}{6}\quad\frac{5}{6}\quad\frac{5}{6}\quad\frac{1}{6}$$

$$= 3\quad\frac{1}{6}\quad\frac{5^2}{6^2}\qquad\frac{75}{216}$$

P(X = 2) = P (two doublets and one non-doublet)

$$= \frac{1}{6}\quad\frac{1}{6}\quad\frac{5}{6}\quad\frac{1}{6}\quad\frac{5}{6}\quad\frac{1}{6}\quad\frac{5}{6}\quad\frac{1}{6}\quad\frac{1}{6}\quad 3\quad\frac{1}{6^2}\quad\frac{5}{6}\qquad\frac{15}{216}$$

and $\quad$ P(X = 3) = P (three doublets)

$$= \frac{1}{6}\quad\frac{1}{6}\quad\frac{1}{6}\quad\frac{1}{216}$$

Thus, the required probability distribution is

X	0	1	2	3
P(X)	$\frac{125}{216}$	$\frac{75}{216}$	$\frac{15}{216}$	$\frac{1}{216}$

Verification Sum of the probabilities

$$\sum_{i=1}^{n} p_i \;=\; \frac{125}{216}\quad\frac{75}{216}\quad\frac{15}{216}\quad\frac{1}{216}$$

$$= \frac{125\quad 75\quad 15\quad 1}{216}\quad\frac{216}{216}\quad 1$$

Example 26 Let X denote the number of hours you study during a randomly selected school day. The probability that X can take the values x, has the following form, where k is some unknown constant.

$$P(X = x) = \begin{cases} 0.1, \text{ if } x = 0 \\ kx, \text{ if } x = 1 \text{ or } 2 \\ k(5 - x), \text{ if } x = 3 \text{ or } 4 \\ 0, \text{ otherwise} \end{cases}$$

(a) Find the value of k.

(b) What is the probability that you study at least two hours ? Exactly two hours? At most two hours?

Solution The probability distribution of X is

X	0	1	2	3	4
P(X)	0.1	k	$2k$	$2k$	k

(a) We know that $\sum_{i=1}^{n} p_i = 1$

Therefore $0.1 + k + 2k + 2k + k = 1$

i.e. $k = 0.15$

(b) P(you study at least two hours) $= P(X \geq 2)$

$= P(X = 2) + P(X = 3) + P(X = 4)$

$= 2k + 2k + k = 5k = 5 \times 0.15 = 0.75$

P(you study exactly two hours) $= P(X = 2)$

$= 2k = 2 \times 0.15 = 0.3$

P(you study at most two hours) $= P(X \leq 2)$

$= P(X = 0) + P(X = 1) + P(X = 2)$

$= 0.1 + k + 2k = 0.1 + 3k = 0.1 + 3 \times 0.15$

$= 0.55$

13.6.2 *Mean of a random variable*

In many problems, it is desirable to describe some feature of the random variable by means of a single number that can be computed from its probability distribution. Few such numbers are mean, median and mode. In this section, we shall discuss mean only. Mean is a measure of location or central tendency in the sense that it roughly locates a *middle* or *average value* of the random variable.

Definition 6 Let X be a random variable whose possible values $x_1, x_2, x_3, ..., x_n$ occur with probabilities $p_1, p_2, p_3,..., p_n$, respectively. The mean of X, denoted by μ, is the number $\sum_{i=1}^{n} x_i p_i$ i.e. the mean of X is the weighted average of the possible values of X, each value being weighted by its probability with which it occurs.

The mean of a random variable X is also called the expectation of X, denoted by E(X).

Thus, $$E(X) = \mu = \sum_{i=1}^{n} x_i p_i = x_1 p_1 + x_2 p_2 + ... + x_n p_n.$$

In other words, the mean or expectation of a random variable X is the sum of the products of all possible values of X by their respective probabilities.

Example 27 Let a pair of dice be thrown and the random variable X be the sum of the numbers that appear on the two dice. Find the mean or expectation of X.

Solution The sample space of the experiment consists of 36 elementary events in the form of ordered pairs (x_i, y_i), where $x_i = 1, 2, 3, 4, 5, 6$ and $y_i = 1, 2, 3, 4, 5, 6$.

The random variable X i.e. the sum of the numbers on the two dice takes the values 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 or 12.

Now $$P(X = 2) = P(\{(1,1)\}) = \frac{1}{36}$$

$$P(X = 3) = P(\{(1,2), (2,1)\}) = \frac{2}{36}$$

$$P(X = 4) = P(\{(1,3), (2,2), (3,1)\}) = \frac{3}{36}$$

$$P(X = 5) = P(\{(1,4), (2,3), (3,2), (4,1)\}) = \frac{4}{36}$$

$$P(X = 6) = P(\{(1,5), (2,4), (3,3), (4,2), (5,1)\}) = \frac{5}{36}$$

$$P(X = 7) = P(\{(1,6), (2,5), (3,4), (4,3), (5,2), (6,1)\}) = \frac{6}{36}$$

$$P(X = 8) = P(\{(2,6), (3,5), (4,4), (5,3), (6,2)\}) = \frac{5}{36}$$

$$P(X = 9) = P(\{(3,6), (4,5), (5,4), (6,3)\}) \quad \frac{4}{36}$$

$$P(X = 10) = P(\{(4,6), (5,5), (6,4)\}) \quad \frac{3}{36}$$

$$P(X = 11) = P(\{(5,6), (6,5)\}) \quad \frac{2}{36}$$

$$P(X = 12) = P(\{(6,6)\}) \quad \frac{1}{36}$$

The probability distribution of X is

X or x_i	2	3	4	5	6	7	8	9	10	11	12
P(X) or p_i	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Therefore,

$$\mu = E(X) = \sum_{i=1}^{n} x_i p_i = 2\times\frac{1}{36} + 3\times\frac{2}{36} + 4\times\frac{3}{36} + 5\times\frac{4}{36}$$

$$6 \quad \frac{5}{36} \quad 7 \quad \frac{6}{36} \quad 8 \quad \frac{5}{36} \quad 9 \quad \frac{4}{36} \quad 10 \quad \frac{3}{36} \quad 11 \quad \frac{2}{36} \quad 12 \quad \frac{1}{36}$$

$$= \frac{2 \quad 6 \quad 12 \quad 20 \quad 30 \quad 42 \quad 40 \quad 36 \quad 30 \quad 22 \quad 12}{36} = 7$$

Thus, the mean of the sum of the numbers that appear on throwing two fair dice is 7.

13.6.3 *Variance of a random variable*

The mean of a random variable does not give us information about the variability in the values of the random variable. In fact, if the variance is small, then the values of the random variable are close to the mean. Also random variables with different probability distributions can have equal means, as shown in the following distributions of X and Y.

X	1	2	3	4
P(X)	$\frac{1}{8}$	$\frac{2}{8}$	$\frac{3}{8}$	$\frac{2}{8}$

Y	–1	0	4	5	6
P(Y)	$\frac{1}{8}$	$\frac{2}{8}$	$\frac{3}{8}$	$\frac{1}{8}$	$\frac{1}{8}$

Clearly $$\text{E(X)} = 1\times\frac{1}{8}+2\times\frac{2}{8}+3\times\frac{3}{8}+4\times\frac{2}{8}=\frac{22}{8}=2.75$$

and $$\text{E(Y)} = -1\times\frac{1}{8}+0\times\frac{2}{8}+4\times\frac{3}{8}+5\times\frac{1}{8}= 6\times\frac{1}{8}=\frac{22}{8}=2.75$$

The variables X and Y are different, however their means are same. It is also easily observable from the diagramatic representation of these distributions (Fig 13.5).

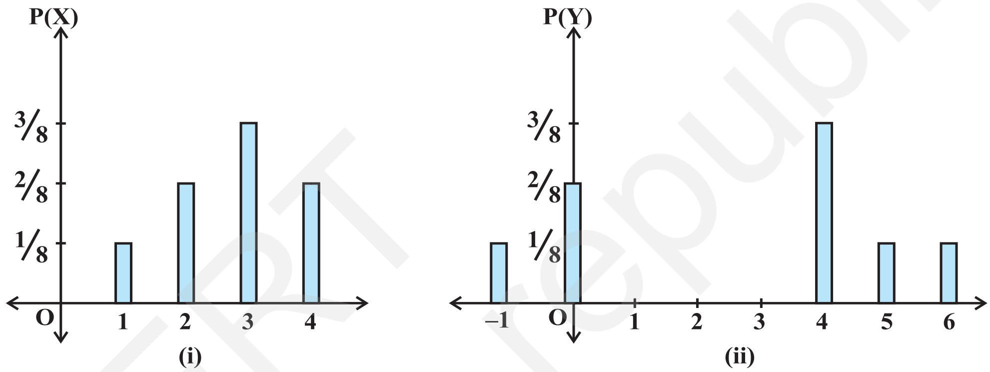


Fig 13.5

To distinguish X from Y, we require a measure of the extent to which the values of the random variables spread out. In Statistics, we have studied that the variance is a measure of the spread or scatter in data. Likewise, the variability or spread in the values of a random variable may be measured by variance.

Definition 7 Let X be a random variable whose possible values $x_1, x_2,...,x_n$ occur with probabilities $p(x_1), p(x_2),..., p(x_n)$ respectively.

Let μ = E (X) be the mean of X. The variance of X, denoted by Var (X) or ${}_x^2$ is defined as

$$\sigma_x^2 = \text{Var}(\text{X}) = \sum_{i=1}^{n}(x_i-\mu)^2\, p(x_i)$$

or equivalently $${}_x^2 = \text{E}(\text{X} - \mu)^2$$

The non-negative number

$$\sigma_x = \sqrt{\text{Var(X)}} = \sqrt{\sum_{i=1}^{n}(x_i - \mu)^2 \, p(x_i)}$$

is called the *standard deviation* of the random variable X.

Another formula to find the variance of a random variable. We know that,

$$\text{Var (X)} = \sum_{i=1}^{n}(x_i - \mu)^2 \, p(x_i)$$

$$= \sum_{i=1}^{n}(x_i^2 + \mu^2 - 2\mu\, x_i)\, p(x_i)$$

$$= \sum_{i=1}^{n} x_i^2 \, p(x_i) + \sum_{i=1}^{n} \mu^2 \, p(x_i) - \sum_{i=1}^{n} 2\mu x_i \, p(x_i)$$

$$= \sum_{i=1}^{n} x_i^2 \, p(x_i) + \mu^2 \sum_{i=1}^{n} p(x_i) - 2\mu \sum_{i=1}^{n} x_i \, p(x_i)$$

$$= \sum_{i=1}^{n} x_i^2 \, p(x_i) + \mu^2 - 2\mu^2 \left[\text{since} \sum_{i=1}^{n} p\,(x_i) = 1 \,\text{and}\, \mu = \sum_{i=1}^{n} x_i \, p(x_i)\right]$$

$$= \sum_{i=1}^{n} x_i^2 \, p(x_i) - \mu^2$$

or $$\text{Var (X)} = \sum_{i=1}^{n} x_i^2 \, p(x_i) - \left(\sum_{i=1}^{n} x_i \, p(x_i)\right)^2$$

or $$\text{Var (X)} = \text{E}(\text{X}^2) - [\text{E}(\text{X})]^2, \text{ where } \text{E}(\text{X}^2) = \sum_{i=1}^{n} x_i^2 \, p(x_i)$$

Example 28 Find the variance of the number obtained on a throw of an unbiased die.

Solution The sample space of the experiment is S = {1, 2, 3, 4, 5, 6}.

Let X denote the number obtained on the throw. Then X is a random variable which can take values 1, 2, 3, 4, 5, or 6.

Also $\quad P(1) = P(2) = P(3) = P(4) = P(5) = P(6) = \frac{1}{6}$

Therefore, the Probability distribution of X is

X	1	2	3	4	5	6
P(X)	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Now $$E(X) = \sum_{i=1}^{n} x_i\, p(x_i)$$

$$= 1\times\frac{1}{6}+2\times\frac{1}{6}+3\times\frac{1}{6}+4\times\frac{1}{6}+5\times\frac{1}{6}+6\times\frac{1}{6}=\frac{21}{6}$$

Also $$E(X^2) = 1^2\times\frac{1}{6}+2^2\times\frac{1}{6}+3^2\times\frac{1}{6}+4^2\times\frac{1}{6}+5^2\times\frac{1}{6}+6^2\times\frac{1}{6}=\frac{91}{6}$$

Thus, $$\text{Var}(X) = E(X^2) - (E(X))^2$$

$$= \frac{91}{6}-\left(\frac{21}{6}\right)^2 = \frac{91}{6} - \frac{441}{36} = \frac{35}{12}$$

Example 29 Two cards are drawn simultaneously (or successively without replacement) from a well shuffled pack of 52 cards. Find the mean, variance and standard deviation of the number of kings.

Solution Let X denote the number of kings in a draw of two cards. X is a random variable which can assume the values 0, 1 or 2.

Now $$P(X = 0) = P(\text{no king}) = \frac{{}^{48}C_2}{{}^{52}C_2} = \frac{\frac{48!}{2!(48-2)!}}{\frac{52!}{2!(52-2)!}} = \frac{48\times 47}{52\times 51} = \frac{188}{221}$$

$$P(X = 1) = P(\text{one king and one non-king}) = \frac{{}^{4}C_1\ {}^{48}C_1}{{}^{52}C_2}$$

$$= \frac{4\times 48\times 2}{52\times 51} = \frac{32}{221}$$

and $\quad P(X = 2) = P \text{ (two kings)} = \dfrac{{}^{4}C_2}{{}^{52}C_2} = \dfrac{4 \times 3}{52 \times 51} = \dfrac{1}{221}$

Thus, the probability distribution of X is

X	0	1	2
P(X)	$\dfrac{188}{221}$	$\dfrac{32}{221}$	$\dfrac{1}{221}$

Now $\qquad$ Mean of $\quad X = E(X) = \sum_{i=1}^{n} x_i \, p(x_i)$

$$= 0 \times \frac{188}{221} + 1 \times \frac{32}{221} + 2 \times \frac{1}{221} = \frac{34}{221}$$

Also $\qquad E(X^2) = \sum_{i=1}^{n} x_i^2 \, p(x_i)$

$$= 0^2 \times \frac{188}{221} + 1^2 \times \frac{32}{221} + 2^2 \times \frac{1}{221} = \frac{36}{221}$$

Now $\qquad Var(X) = E(X^2) - [E(X)]^2$

$$= \frac{36}{221} - \left(\frac{34}{221}\right)^2 = \frac{6800}{(221)^2}$$

Therefore $\qquad \sigma_x = \sqrt{Var(X)} = \dfrac{\sqrt{6800}}{221} = 0.37$

EXERCISE 13.4

1. State which of the following are not the probability distributions of a random variable. Give reasons for your answer.

(i)

X	0	1	2
P(X)	0.4	0.4	0.2

(ii)

X	0	1	2	3	4
P(X)	0.1	0.5	0.2	– 0.1	0.3

(iii)

Y	– 1	0	1
P(Y)	0.6	0.1	0.2

(iv)

Z	3	2	1	0	–1
P(Z)	0.3	0.2	0.4	0.1	0.05

2. An urn contains 5 red and 2 black balls. Two balls are randomly drawn. Let X represent the number of black balls. What are the possible values of X? Is X a random variable ?

3. Let X represent the difference between the number of heads and the number of tails obtained when a coin is tossed 6 times. What are possible values of X?

4. Find the probability distribution of

(i) number of heads in two tosses of a coin.

(ii) number of tails in the simultaneous tosses of three coins.

(iii) number of heads in four tosses of a coin.

5. Find the probability distribution of the number of successes in two tosses of a die, where a success is defined as

(i) number greater than 4

(ii) six appears on at least one die

6. From a lot of 30 bulbs which include 6 defectives, a sample of 4 bulbs is drawn at random with replacement. Find the probability distribution of the number of defective bulbs.

7. A coin is biased so that the head is 3 times as likely to occur as tail. If the coin is tossed twice, find the probability distribution of number of tails.

8. A random variable X has the following probability distribution:

X	0	1	2	3	4	5	6	7
P(X)	0	k	$2k$	$2k$	$3k$	k^2	$2k^2$	$7k^2+k$

Determine

(i) k (ii) P(X < 3)

(iii) P(X > 6) (iv) P(0 < X < 3)

9. The random variable X has a probability distribution P(X) of the following form, where k is some number :

$$P(X) = \begin{cases} k, & \text{if } x=0 \\ 2k, & \text{if } x=1 \\ 3k, & \text{if } x=2 \\ 0, & \text{otherwise} \end{cases}$$

(a) Determine the value of k.

(b) Find P $(X < 2)$, P $(X \leq 2)$, P$(X \geq 2)$.

10. Find the mean number of heads in three tosses of a fair coin.

11. Two dice are thrown simultaneously. If X denotes the number of sixes, find the expectation of X.

12. Two numbers are selected at random (without replacement) from the first six positive integers. Let X denote the larger of the two numbers obtained. Find E(X).

13. Let X denote the sum of the numbers obtained when two fair dice are rolled. Find the variance and standard deviation of X.

14. A class has 15 students whose ages are 14, 17, 15, 14, 21, 17, 19, 20, 16, 18, 20, 17, 16, 19 and 20 years. One student is selected in such a manner that each has the same chance of being chosen and the age X of the selected student is recorded. What is the probability distribution of the random variable X? Find mean, variance and standard deviation of X.

15. In a meeting, 70% of the members favour and 30% oppose a certain proposal. A member is selected at random and we take X = 0 if he opposed, and X = 1 if he is in favour. Find E(X) and Var (X).

Choose the correct answer in each of the following:

16. The mean of the numbers obtained on throwing a die having written 1 on three faces, 2 on two faces and 5 on one face is

(A) 1 (B) 2 (C) 5 (D) $\frac{8}{3}$

17. Suppose that two cards are drawn at random from a deck of cards. Let X be the number of aces obtained. Then the value of E(X) is

(A) $\frac{37}{221}$ (B) $\frac{5}{13}$ (C) $\frac{1}{13}$ (D) $\frac{2}{13}$

13.7 Bernoulli Trials and Binomial Distribution

13.7.1 *Bernoulli trials*

Many experiments are dichotomous in nature. For example, a tossed coin shows a 'head' or 'tail', a manufactured item can be 'defective' or 'non-defective', the response to a question might be 'yes' or 'no', an egg has 'hatched' or 'not hatched', the decision is 'yes' or 'no' etc. In such cases, it is customary to call one of the outcomes a 'success' and the other 'not success' or 'failure'. For example, in tossing a coin, if the occurrence of the head is considered a success, then occurrence of tail is a failure.

Each time we toss a coin or roll a die or perform any other experiment, we call it a trial. If a coin is tossed, say, 4 times, the number of trials is 4, each having exactly two outcomes, namely, success or failure. The outcome of any trial is independent of the outcome of any other trial. In each of such trials, the probability of success or failure remains constant. Such independent trials which have only two outcomes usually referred as 'success' or 'failure' are called *Bernoulli trials*.

Definition 8 Trials of a random experiment are called Bernoulli trials, if they satisfy the following conditions :

(i) There should be a finite number of trials.

(ii) The trials should be independent.

(iii) Each trial has exactly two outcomes : success or failure.

(iv) The probability of success remains the same in each trial.

For example, throwing a die 50 times is a case of 50 Bernoulli trials, in which each trial results in success (say an even number) or failure (an odd number) and the probability of success (p) is same for all 50 throws. Obviously, the successive throws of the die are independent experiments. If the die is fair and have six numbers 1 to 6 written on six faces, then $p = \frac{1}{2}$ and $q = 1 - p = \frac{1}{2}$ = probability of failure.

Example 30 Six balls are drawn successively from an urn containing 7 red and 9 black balls. Tell whether or not the trials of drawing balls are Bernoulli trials when after each draw the ball drawn is

(i) replaced (ii) not replaced in the urn.

Solution

(i) The number of trials is finite. When the drawing is done with replacement, the probability of success (say, red ball) is $p = \frac{7}{16}$ which is same for all six trials (draws). Hence, the drawing of balls with replacements are Bernoulli trials.

(ii) When the drawing is done without replacement, the probability of success (i.e., red ball) in first trial is $\frac{7}{16}$, in 2nd trial is $\frac{6}{15}$ if the first ball drawn is red or $\frac{7}{15}$ if the first ball drawn is black and so on. Clearly, the probability of success is not same for all trials, hence the trials are not Bernoulli trials.

13.7.2 *Binomial distribution*

Consider the experiment of tossing a coin in which each trial results in success (say, heads) or failure (tails). Let S and F denote respectively success and failure in each trial. Suppose we are interested in finding the ways in which we have one success in six trials.

Clearly, six different cases are there as listed below:

SFFFFF, FSFFFF, FFSFFF, FFFSFF, FFFFSF, FFFFFS.

Similarly, two successes and four failures can have $\frac{6!}{4!\ \ 2!}$ combinations. It will be lengthy job to list all of these ways. Therefore, calculation of probabilities of 0, 1, 2,..., n number of successes may be lengthy and time consuming. To avoid the lengthy calculations and listing of all the possible cases, for the probabilities of number of successes in n-Bernoulli trials, a formula is derived. For this purpose, let us take the experiment made up of three Bernoulli trials with probabilities p and $q = 1 - p$ for success and failure respectively in each trial. The sample space of the experiment is the set

$$S = \{SSS, SSF, SFS, FSS, SFF, FSF, FFS, FFF\}$$

The number of successes is a random variable X and can take values 0, 1, 2, or 3. The probability distribution of the number of successes is as below :

$$\begin{aligned}
P(X = 0) &= P(\text{no success}) \\
&= P(\{FFF\}) = P(F)\ P(F)\ P(F) \\
&= q\ .\ q\ .\ q = q^3 \text{ since the trials are independent} \\
P(X = 1) &= P(\text{one successes}) \\
&= P(\{SFF, FSF, FFS\}) \\
&= P(\{SFF\}) + P(\{FSF\}) + P(\{FFS\}) \\
&= P(S)\ P(F)\ P(F) + P(F)\ P(S)\ P(F) + P(F)\ P(F)\ P(S) \\
&= p.q.q + q.p.q + q.q.p = 3pq^2 \\
P(X = 2) &= P\ (\text{two successes}) \\
&= P(\{SSF, SFS, FSS\}) \\
&= P(\{SSF\}) + P\ (\{SFS\}) + P(\{FSS\})
\end{aligned}$$

$$= \mathrm{P(S)\ P(S)\ P(F) + P(S)\ P(F)\ P(S) + P(F)\ P(S)\ P(S)}$$

$$= p.p.q. + p.q.p + q.p.p = 3p^2q$$

and $\quad \mathrm{P(X = 3)} = \mathrm{P(three\ success)} = \mathrm{P(\{SSS\})}$

$$= \mathrm{P(S)\ .\ P(S)\ .\ P(S)} = p^3$$

Thus, the probability distribution of X is

X	0	1	2	3
P(X)	q^3	$3q^2p$	$3qp^2$	p^3

Also, the binominal expansion of $(q + p)^3$ is

$$q^3 + 3q^2 p + 3qp^2 + p^3$$

Note that the probabilities of 0, 1, 2 or 3 successes are respectively the 1st, 2nd, 3rd and 4th term in the expansion of $(q + p)^3$.

Also, since $q + p = 1$, it follows that the sum of these probabilities, as expected, is 1.

Thus, we may conclude that in an experiment of n-Bernoulli trials, the probabilities of 0, 1, 2,..., n successes can be obtained as 1st, 2nd,...,$(n + 1)^{\text{th}}$ terms in the expansion of $(q + p)^n$. To prove this assertion (result), let us find the probability of x-successes in an experiment of n-Bernoulli trials.

Clearly, in case of x successes (S), there will be $(n – x)$ failures (F).

Now, x successes (S) and $(n – x)$ failures (F) can be obtained in $\dfrac{n!}{x!(n-x)!}$ ways.

In each of these ways, the probability of x successes and $(n - x)$ failures is

$$= \mathrm{P}(x \text{ successes}) \ .\ \mathrm{P}(n–x) \text{ failures is}$$

$$= \underbrace{\mathrm{P(S).P(S)...P(S)}}_{x\,\text{times}} \quad \underbrace{\mathrm{P(F).P(F)...P(F)}}_{(n-x)\,\text{times}} = p^x\, q^{n-x}$$

Thus, the probability of x successes in n-Bernoulli trials is $\dfrac{n!}{x!(n-x)!}p^x\, q^{n-x}$

or ${}^nC_x\, p^x\ q^{n-x}$

Thus $\qquad \mathrm{P}(x \text{ successes}) = {}^nC_x\, p^x\, q^{n-x}, \quad x = 0, 1, 2,...,n.\ (q = 1 – p)$

Clearly, P(x successes), i.e. ${}^nC_x\, p^x\, q^{n-x}$ is the $(x + 1)^{\text{th}}$ term in the binomial expansion of $(q + p)^n$.

Thus, the probability distribution of number of successes in an experiment consisting of n Bernoulli trials may be obtained by the binomial expansion of $(q + p)^n$. Hence, this distribution of number of successes X can be written as

X	0	1	2	…	x	…	n
P(X)	${}^nC_0\, q^n$	${}^nC_1\, q^{n-1}p^1$	${}^nC_2\, q^{n-2}p^2$		${}^nC_x\, q^{n-x}p^x$		${}^nC_n\, p^n$

The above probability distribution is known as *binomial distribution* with parameters n and p, because for given values of n and p, we can find the complete probability distribution.

The probability of x successes P(X = x) is also denoted by P(x) and is given by

$$P(x) = {}^nC_x q^{n-x}p^x, \quad x = 0, 1, ..., n. \ (q = 1 - p)$$

This P(x) is called the *probability function* of the binomial distribution.

A binomial distribution with n-Bernoulli trials and probability of success in each trial as p, is denoted by B(n, p).

Let us now take up some examples.

Example 31 If a fair coin is tossed 10 times, find the probability of

(i) exactly six heads

(ii) at least six heads

(iii) at most six heads

Solution The repeated tosses of a coin are Bernoulli trials. Let X denote the number of heads in an experiment of 10 trials.

Clearly, X has the binomial distribution with $n = 10$ and $p = \frac{1}{2}$

Therefore $\quad P(X = x) = {}^nC_x q^{n-x}p^x, x = 0, 1, 2, ..., n$

Here $\quad n = 10,\ p\ \frac{1}{2}, q = 1 - p = \frac{1}{2}$

Therefore $\quad P(X = x) = {}^{10}C_x \left(\frac{1}{2}\right)^{10-x} \left(\frac{1}{2}\right)^x = {}^{10}C_x \left(\frac{1}{2}\right)^{10}$

Now (i) $P(X = 6) = {}^{10}C_6 \left(\frac{1}{2}\right)^{10} = \frac{10!}{6! \times 4!} \frac{1}{2^{10}} = \frac{105}{512}$

(ii) P(at least six heads) = P(X ≥ 6)

$$= P(X = 6) + P(X = 7) + P(X = 8) + P(X = 9) + P(X = 10)$$

$$= {}^{10}C_6\left(\frac{1}{2}\right)^{10} + {}^{10}C_7\left(\frac{1}{2}\right)^{10} + {}^{10}C_8\left(\frac{1}{2}\right)^{10} + {}^{10}C_9\left(\frac{1}{2}\right)^{10} + {}^{10}C_{10}\left(\frac{1}{2}\right)^{10}$$

$$= \frac{10!}{6!\quad 4!} \quad \frac{10!}{7!\quad 3!} \quad \frac{10!}{8!\quad 2!} \quad \frac{10!}{9!\quad 1!} \quad \frac{10!}{10!} \quad \frac{1}{2^{10}} = \frac{193}{512}$$

(iii) P(at most six heads) = P(X ≤ 6)

$$= P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3)$$
$$+ P(X = 4) + P(X = 5) + P(X = 6)$$

$$= \left(\frac{1}{2}\right)^{10} + {}^{10}C_1\left(\frac{1}{2}\right)^{10} + {}^{10}C_2\left(\frac{1}{2}\right)^{10} + {}^{10}C_3\left(\frac{1}{2}\right)^{10}$$

$$+ {}^{10}C_4\left(\frac{1}{2}\right)^{10} + {}^{10}C_5\left(\frac{1}{2}\right)^{10} + {}^{10}C_6\left(\frac{1}{2}\right)^{10}$$

$$= \frac{848}{1024} = \frac{53}{64}$$

Example 32 Ten eggs are drawn successively with replacement from a lot containing 10% defective eggs. Find the probability that there is at least one defective egg.

Solution Let X denote the number of defective eggs in the 10 eggs drawn. Since the drawing is done with replacement, the trials are Bernoulli trials. Clearly, X has the binomial distribution with $n = 10$ and $p \quad \frac{10}{100} \quad \frac{1}{10}$.

Therefore $$q = 1 - p = \frac{9}{10}$$

Now P(at least one defective egg) = P(X ≥ 1) = 1 – P (X = 0)

$$= 1 - {}^{10}C_0\left(\frac{9}{10}\right)^{10} = 1 - \frac{9^{10}}{10^{10}}$$

EXERCISE 13.5

1. A die is thrown 6 times. If 'getting an odd number' is a success, what is the probability of
   (i) 5 successes? (ii) at least 5 successes?
   (iii) at most 5 successes?

2. A pair of dice is thrown 4 times. If getting a doublet is considered a success, find the probability of two successes.

3. There are 5% defective items in a large bulk of items. What is the probability that a sample of 10 items will include not more than one defective item?

4. Five cards are drawn successively with replacement from a well-shuffled deck of 52 cards. What is the probability that

(i) all the five cards are spades?

(ii) only 3 cards are spades?

(iii) none is a spade?

5. The probability that a bulb produced by a factory will fuse after 150 days of use is 0.05. Find the probability that out of 5 such bulbs

(i) none

(ii) not more than one

(iii) more than one

(iv) at least one

will fuse after 150 days of use.

6. A bag consists of 10 balls each marked with one of the digits 0 to 9. If four balls are drawn successively with replacement from the bag, what is the probability that none is marked with the digit 0?

7. In an examination, 20 questions of true-false type are asked. Suppose a student tosses a fair coin to determine his answer to each question. If the coin falls heads, he answers 'true'; if it falls tails, he answers 'false'. Find the probability that he answers at least 12 questions correctly.

8. Suppose X has a binomial distribution $B\left(6, \frac{1}{2}\right)$. Show that $X = 3$ is the most likely outcome.

(Hint : $P(X = 3)$ is the maximum among all $P(x_i)$, $x_i = 0,1,2,3,4,5,6$)

9. On a multiple choice examination with three possible answers for each of the five questions, what is the probability that a candidate would get four or more correct answers just by guessing ?

10. A person buys a lottery ticket in 50 lotteries, in each of which his chance of winning a prize is $\frac{1}{100}$. What is the probability that he will win a prize

(a) at least once (b) exactly once (c) at least twice?

11. Find the probability of getting 5 exactly twice in 7 throws of a die.

12. Find the probability of throwing at most 2 sixes in 6 throws of a single die.

13. It is known that 10% of certain articles manufactured are defective. What is the probability that in a random sample of 12 such articles, 9 are defective?

In each of the following, choose the correct answer:

14. In a box containing 100 bulbs, 10 are defective. The probability that out of a sample of 5 bulbs, none is defective is

(A) 10^{-1} (B) $\left(\frac{1}{2}\right)^5$ (C) $\left(\frac{9}{10}\right)^5$ (D) $\frac{9}{10}$

15. The probability that a student is not a swimmer is $\frac{1}{5}\cdot$ Then the probability that out of five students, four are swimmers is

(A) ${}^5C_4\left(\frac{4}{5}\right)^4\frac{1}{5}$ (B) $\left(\frac{4}{5}\right)^4\frac{1}{5}$

(C) ${}^5C_1\frac{1}{5}\left(\frac{4}{5}\right)^4$ (D) None of these

Miscellaneous Examples

Example 33 Coloured balls are distributed in four boxes as shown in the following table:

Box	Colour			
	Black	**White**	**Red**	**Blue**
I	3	4	5	6
II	2	2	2	2
III	1	2	3	1
IV	4	3	1	5

A box is selected at random and then a ball is randomly drawn from the selected box. The colour of the ball is black, what is the probability that ball drawn is from the box III?

Solution Let A, E_1, E_2, E_3 and E_4 be the events as defined below :

A : a black ball is selected E_1 : box I is selected

E_2 : box II is selected E_3 : box III is selected

E_4 : box IV is selected

Since the boxes are chosen at random,

Therefore $\quad P(E_1) = P(E_2) = P(E_3) = P(E_4) = \frac{1}{4}$

Also $\quad P(A|E_1) = \frac{3}{18}$, $P(A|E_2) = \frac{2}{8}$, $P(A|E_3) = \frac{1}{7}$ and $P(A|E_4) = \frac{4}{13}$

P(box III is selected, given that the drawn ball is black) = $P(E_3|A)$. By Bayes' theorem,

$$P(E_3|A) = \frac{P(E_3)\ P(A|E_3)}{P(E_1)P(A|E_1)\quad P(E_2)P(A|E_2)+P(E_3)P(A|E_3)\quad P(E_4)P(A|E_4)}$$

$$= \frac{\frac{1}{4}\times\frac{1}{7}}{\frac{1}{4}\times\frac{3}{18}+\frac{1}{4}\times\frac{1}{4}+\frac{1}{4}\times\frac{1}{7}+\frac{1}{4}\times\frac{4}{13}} = 0.165$$

Example 34 Find the mean of the Binomial distribution $B\ \ 4, \frac{1}{3}$.

Solution Let X be the random variable whose probability distribution is $B\ \ 4, \frac{1}{3}$.

Here $\quad n = 4,\ p = \frac{1}{3}$ and $q = 1 - \frac{1}{3} = \frac{2}{3}$

We know that $\quad P(X = x) = {}^4C_x \left(\frac{2}{3}\right)^{4-x}\left(\frac{1}{3}\right)^x$, $x = 0, 1, 2, 3, 4.$

i.e. the distribution of X is

x_i	$P(x_i)$	$x_i\ P(x_i)$
0	${}^4C_0\ \frac{2}{3}^{\ 4}$	0
1	${}^4C_1\ \frac{2}{3}^{\ 3}\ \frac{1}{3}$	${}^4C_1\ \frac{2}{3}^{\ 3}\ \frac{1}{3}$

2	${}^4C_2\ \frac{2}{3}^{\,2}\ \frac{1}{3}^{\,2}$	$2\ {}^4C_2\ \frac{2}{3}^{\,2}\ \frac{1}{3}^{\,2}$
3	${}^4C_3\ \frac{2}{3}\ \frac{1}{3}^{\,3}$	$3\ {}^4C_3\ \frac{2}{3}\ \frac{1}{3}^{\,3}$
4	${}^4C_4\ \frac{1}{3}^{\,4}$	$4\ {}^4C_4\ \frac{1}{3}^{\,4}$

Now Mean (μ) $= \sum_{i=1}^{4} x_i\ p(x_i)$

$$= 0 + {}^4C_1\left(\frac{2}{3}\right)^3\left(\frac{1}{3}\right) + 2\cdot{}^4C_2\left(\frac{2}{3}\right)^2\left(\frac{1}{3}\right)^2 + 3\ {}^4C_3\ \frac{2}{3}\ \frac{1}{3}^{\,3}\quad 4\ {}^4C_4\ \frac{1}{3}^{\,4}$$

$$= 4\times\frac{2^3}{3^4} + 2\times 6\times\frac{2^2}{3^4} + 3\times 4\times\frac{2}{3^4} + 4\times 1\times\frac{1}{3^4}$$

$$= \frac{32+48+24+4}{3^4} = \frac{108}{81} = \frac{4}{3}$$

Example 35 The probability of a shooter hitting a target is $\frac{3}{4}$. How many minimum number of times must he/she fire so that the probability of hitting the target at least once is more than 0.99?

Solution Let the shooter fire n times. Obviously, n fires are n Bernoulli trials. In each trial, p = probability of hitting the target $= \frac{3}{4}$ and q = probability of not hitting the target $= \frac{1}{4}$. Then $P(X = x) = {}^nC_x\, q^{n-x}\, p^x = {}^nC_x\left(\frac{1}{4}\right)^{n-x}\left(\frac{3}{4}\right)^x = {}^nC_x\,\frac{3^x}{4^n}$.

Now, given that,

P(hitting the target at least once) > 0.99

i.e. $\qquad P(x \geq 1) > 0.99$

Therefore, $\quad 1 - \text{P}(x = 0) > 0.99$

or $\quad 1 - {}^n\text{C}_0 \dfrac{1}{4^n} > 0.99$

or $\quad {}^n\text{C}_0 \dfrac{1}{4^n} \quad 0.01 \text{ i.e. } \dfrac{1}{4^n} < 0.01$

or $\quad 4^n > \dfrac{1}{0.01} = 100 \qquad \ldots (1)$

The minimum value of n to satisfy the inequality (1) is 4.

Thus, the shooter must fire 4 times.

Example 36 A and B throw a die alternatively till one of them gets a '6' and wins the game. Find their respective probabilities of winning, if A starts first.

Solution Let S denote the success (getting a '6') and F denote the failure (not getting a '6').

Thus, $\quad \text{P(S)} = \dfrac{1}{6}, \ \text{P(F)} = \dfrac{5}{6}$

$$\text{P(A wins in the first throw)} = \text{P(S)} = \frac{1}{6}$$

A gets the third throw, when the first throw by A and second throw by B result into failures.

Therefore, $\quad \text{P(A wins in the 3rd throw)} = \text{P(FFS)} = \text{P(F)P(F)P(S)} = \dfrac{5}{6} \ \dfrac{5}{6} \ \dfrac{1}{6}$

$$= \left(\frac{5}{6}\right)^2 \times \frac{1}{6}$$

$$\text{P(A wins in the 5th throw)} = \text{P (FFFFS)} \quad \frac{5}{6}^{4} \ \frac{1}{6} \text{ and so on.}$$

Hence, $\quad \text{P(A wins)} = \dfrac{1}{6} \quad \dfrac{5}{6}^{2} \ \dfrac{1}{6} \quad \dfrac{5}{6}^{4} \ \dfrac{1}{6} \quad \ldots$

$$= \frac{\frac{1}{6}}{1 - \frac{25}{36}} = \frac{6}{11}$$

$$\text{P(B wins)} = 1 - \text{P (A wins)} = 1 \quad \frac{6}{11} \quad \frac{5}{11}$$

Remark If $a + ar + ar^2 + ... + ar^{n-1} + ...$, where $|r| < 1$, then sum of this infinite G.P. is given by $\frac{a}{1-r}\cdot$ (Refer A.1.3 of Class XI Text book).

Example 37 If a machine is correctly set up, it produces 90% acceptable items. If it is incorrectly set up, it produces only 40% acceptable items. Past experience shows that 80% of the set ups are correctly done. If after a certain set up, the machine produces 2 acceptable items, find the probability that the machine is correctly setup.

Solution Let A be the event that the machine produces 2 acceptable items.

Also let B_1 represent the event of correct set up and B_2 represent the event of incorrect setup.

Now $\quad P(B_1) = 0.8,\ P(B_2) = 0.2$

$$P(A|B_1) = 0.9 \times 0.9 \ \text{ and } P(A|B_2) = 0.4 \times 0.4$$

Therefore
$$P(B_1|A) = \frac{P(B_1)\,P(A|B_1)}{P(B_1)\,P(A|B_1) + P(B_2)\,P(A|B_2)}$$

$$= \frac{0.8 \times 0.9 \times 0.9}{0.8 \times 0.9 \times 0.9 + 0.2 \times 0.4 \times 0.4} = \frac{648}{680} = 0.95$$

Miscellaneous Exercise on Chapter 13

1. A and B are two events such that P (A) ≠ 0. Find P(B|A), if
   (i) A is a subset of B (ii) $A \cap B = \phi$

2. A couple has two children,
   (i) Find the probability that both children are males, if it is known that at least one of the children is male.
   (ii) Find the probability that both children are females, if it is known that the elder child is a female.

3. Suppose that 5% of men and 0.25% of women have grey hair. A grey haired person is selected at random. What is the probability of this person being male? Assume that there are equal number of males and females.

4. Suppose that 90% of people are right-handed. What is the probability that at most 6 of a random sample of 10 people are right-handed?

5. An urn contains 25 balls of which 10 balls bear a mark 'X' and the remaining 15 bear a mark 'Y'. A ball is drawn at random from the urn, its mark is noted down and it is replaced. If 6 balls are drawn in this way, find the probability that
   (i) all will bear 'X' mark.
   (ii) not more than 2 will bear 'Y' mark.
   (iii) at least one ball will bear 'Y' mark.
   (iv) the number of balls with 'X' mark and 'Y' mark will be equal.
6. In a hurdle race, a player has to cross 10 hurdles. The probability that he will clear each hurdle is $\frac{5}{6}$. What is the probability that he will knock down fewer than 2 hurdles?
7. A die is thrown again and again until three sixes are obtained. Find the probability of obtaining the third six in the sixth throw of the die.
8. If a leap year is selected at random, what is the chance that it will contain 53 tuesdays?
9. An experiment succeeds twice as often as it fails. Find the probability that in the next six trials, there will be atleast 4 successes.
10. How many times must a man toss a fair coin so that the probability of having at least one head is more than 90%?
11. In a game, a man wins a rupee for a six and loses a rupee for any other number when a fair die is thrown. The man decided to throw a die thrice but to quit as and when he gets a six. Find the expected value of the amount he wins / loses.
12. Suppose we have four boxes A,B,C and D containing coloured marbles as given below:

Box	Marble colour		
	Red	White	Black
A	1	6	3
B	6	2	2
C	8	1	1
D	0	6	4

One of the boxes has been selected at random and a single marble is drawn from it. If the marble is red, what is the probability that it was drawn from box A?, box B?, box C?

13. Assume that the chances of a patient having a heart attack is 40%. It is also assumed that a meditation and yoga course reduce the risk of heart attack by 30% and prescription of certain drug reduces its chances by 25%. At a time a patient can choose any one of the two options with equal probabilities. It is given that after going through one of the two options the patient selected at random suffers a heart attack. Find the probability that the patient followed a course of meditation and yoga?

14. If each element of a second order determinant is either zero or one, what is the probability that the value of the determinant is positive? (Assume that the individual entries of the determinant are chosen independently, each value being assumed with probability $\frac{1}{2}$).

15. An electronic assembly consists of two subsystems, say, A and B. From previous testing procedures, the following probabilities are assumed to be known:

$$P(\text{A fails}) = 0.2$$
$$P(\text{B fails alone}) = 0.15$$
$$P(\text{A and B fail}) = 0.15$$

Evaluate the following probabilities

(i) $P(\text{A fails}|\text{B has failed})$ (ii) $P(\text{A fails alone})$

16. Bag I contains 3 red and 4 black balls and Bag II contains 4 red and 5 black balls. One ball is transferred from Bag I to Bag II and then a ball is drawn from Bag II. The ball so drawn is found to be red in colour. Find the probability that the transferred ball is black.

Choose the correct answer in each of the following:

17. If A and B are two events such that $P(A) \neq 0$ and $P(B \mid A) = 1$, then

(A) $A \subset B$ (B) $B \subset A$ (C) $B = \phi$ (D) $A = \phi$

18. If $P(A|B) > P(A)$, then which of the following is correct :

(A) $P(B|A) < P(B)$ (B) $P(A \cap B) < P(A) \cdot P(B)$

(C) $P(B|A) > P(B)$ (D) $P(B|A) = P(B)$

19. If A and B are any two events such that $P(A) + P(B) – P(A \text{ and } B) = P(A)$, then

(A) $P(B|A) = 1$ (B) $P(A|B) = 1$

(C) $P(B|A) = 0$ (D) $P(A|B) = 0$

Summary

The salient features of the chapter are –

- The conditional probability of an event E, given the occurrence of the event F is given by $P(E|F) = \frac{P(E \cap F)}{P(F)}$, $P(F) \neq 0$
- $0 \le P(E|F) \le 1$, $P(E'|F) = 1 - P(E|F)$
  $P((E \cup F)|G) = P(E|G) + P(F|G) - P((E \cap F)|G)$
- $P(E \cap F) = P(E)\, P(F|E)$, $P(E) \neq 0$
  $P(E \cap F) = P(F)\, P(E|F)$, $P(F) \neq 0$
- If E and F are independent, then
  $P(E \cap F) = P(E)\, P(F)$
  $P(E|F) = P(E)$, $P(F) \neq 0$
  $P(F|E) = P(F)$, $P(E) \neq 0$
- **Theorem of total probability**
  Let $\{E_1, E_2, ..., E_n)$ be a partition of a sample space and suppose that each of $E_1, E_2, ..., E_n$ has nonzero probability. Let A be any event associated with S, then
  $P(A) = P(E_1)\, P(A|E_1) + P(E_2)\, P(A|E_2) + ... + P(E_n)\, P(A|E_n)$
- **Bayes' theorem** If $E_1, E_2, ..., E_n$ are events which constitute a partition of sample space S, i.e. $E_1, E_2, ..., E_n$ are pairwise disjoint and $E_1 \cup E_2 \cup ... \cup E_n = S$ and A be any event with nonzero probability, then

$$P(E_i|A) = \frac{P(E_i)P(A|E_i)}{\sum_{j=1}^{n} P(E_j)P(A|E_j)}$$

- A random variable is a real valued function whose domain is the sample space of a random experiment.
- The probability distribution of a random variable X is the system of numbers

X	:	x_1	x_2	...	x_n
P(X)	:	p_1	p_2	...	p_n

where, $p_i > 0, \sum_{i=1}^{n} p_i = 1, i = 1, 2, ..., n$

- Let X be a random variable whose possible values $x_1, x_2, x_3, ..., x_n$ occur with probabilities $p_1, p_2, p_3, ... p_n$ respectively. The mean of X, denoted by μ, is the number $\sum_{i=1}^{n} x_i p_i$.

  The mean of a random variable X is also called the expectation of X, denoted by E (X).
- Let X be a random variable whose possible values $x_1, x_2, ..., x_n$ occur with probabilities $p(x_1), p(x_2), ..., p(x_n)$ respectively.

  Let μ = E(X) be the mean of X. The variance of X, denoted by Var (X) or σ_x^2, is defined as $\sigma_x^2 = \text{Var}(X) = \sum_{i=1}^{n} (x_i - \mu)^2 p(x_i)$

  or equivalently $\sigma_x^2 = E(X - \mu)^2$

  The non-negative number

$$\sigma_x = \sqrt{\text{Var}(X)} = \sqrt{\sum_{i=1}^{n} (x_i - \mu)^2 p(x_i)}$$

  is called the standard deviation of the random variable X.
- Var (X) = E (X^2) – $[E(X)]^2$
- Trials of a random experiment are called Bernoulli trials, if they satisfy the following conditions :
  - (i) There should be a finite number of trials.
  - (ii) The trials should be independent.
  - (iii) Each trial has exactly two outcomes : success or failure.
  - (iv) The probability of success remains the same in each trial.

  For Binomial distribution B (n, p), P $(X = x) = {}^nC_x \, q^{n-x} p^x$, $x = 0, 1, ..., n$ $(q = 1 - p)$

Historical Note

The earliest indication on measurement of chances in game of dice appeared in 1477 in a commentary on Dante's Divine Comedy. A treatise on gambling named *liber de Ludo Alcae*, by Geronimo Carden (1501-1576) was published posthumously in 1663. In this treatise, he gives the number of favourable cases for each event when two dice are thrown.

Galileo (1564-1642) gave casual remarks concerning the correct evaluation of chance in a game of three dice. Galileo analysed that when three dice are thrown, the sum of the number that appear is more likely to be 10 than the sum 9, because the number of cases favourable to 10 are more than the number of cases for the appearance of number 9.

Apart from these early contributions, it is generally acknowledged that the true origin of the science of probability lies in the correspondence between two great men of the seventeenth century, Pascal (1623-1662) and Pierre de Fermat (1601-1665). A French gambler, Chevalier de Metre asked Pascal to explain some seeming contradiction between his theoretical reasoning and the observation gathered from gambling. In a series of letters written around 1654, Pascal and Fermat laid the first foundation of science of probability. Pascal solved the problem in algebraic manner while Fermat used the method of combinations.

Great Dutch Scientist, Huygens (1629-1695), became acquainted with the content of the correspondence between Pascal and Fermat and published a first book on probability, "*De Ratiociniis in Ludo Aleae*" containing solution of many interesting rather than difficult problems on probability in games of chances.

The next great work on probability theory is by Jacob Bernoulli (1654-1705), in the form of a great book, "*Ars Conjectendi*" published posthumously in 1713 by his nephew, Nicholes Bernoulli. To him is due the discovery of one of the most important probability distribution known as Binomial distribution. The next remarkable work on probability lies in 1993. A. N. Kolmogorov (1903-1987) is credited with the axiomatic theory of probability. His book, 'Foundations of probability' published in 1933, introduces probability as a set function and is considered a 'classic!'.