Toppersnotes®
Unleash the topper in you

CSIR NET

Mathematical Science

Council of Scientific & Industrial Research (CSIR)

Volume - 4

SIERRA
INNOVATIONS

INDEX

1
CHAPTER
Statistics

PROBABILITY

1.1 Outcomes and Events

We consider experiments, which comprise: a collection of distinguishable outcomes, which are termed elementary events, and typically denoted by Ω and a collection of sets of possible outcomes to which we might wish to assign probabilities, A, the event.

In order to obtain a sensible theory of probability, we require that our collection of events A is an algebra over, i.e. it must possess the following properties

(i.) $\Omega \in A$

(ii.) If $A \in A$, then $\underline{A} \in A$

(iii.) If A_1 and $A_2 \in A$, then $A_1 \cup A_2 \in A$.

In the case of finite Ω, we might note that the collection of all subsets of Ω necessarily satisfies the above properties and by using this default choice of algebra , we can assign probabilities to any possible combination of elementary events.

Proposition 1.1: If A is an algebra, then $\phi \in A$.

Proposition 1.2: If A_1 and $A_2 \in A$, then $A_1 \cap A_2 \in A$ for any algebra A.

Proposition 1.3: If A is an algebra and $A_1, A_2, \ldots, A_n \in A$, then $\bigcap_{i=1}^{n} A_i \in A$.

1.2 Probability Functions/Measures

Let Ω denote the sample space and A denote a collection of events assumed to be a σ-algebra.

Definition 1.1: (Probability Function): A probability function $P[.] is$ a set function with domain A (a σ-algebra of events) and range $[0,1]$, i.e., $P: A \rightarrow [0,1]$, which satisfies the following axioms

(i.) $P[A] \geqslant 0$ for every $A \in A$

(ii.) $P[\Omega] = 1$

(iii.) If $A_1, A_2, \ldots$ is a sequence of mutually exclusive events (i.e. $A_i \cap A_j = \phi$ for any $i \neq j$) in A and if $\bigcup_{i=1}^{\infty} A_i \in A$, then $P[\,_i \bigcup_1^{\infty} A_i] = \sum_{i=1}^{\infty} P\ [A_i]$

Properties of P[.

A remarkably rich theory emerges from these three axioms (together, of course, with those of set theory). Indeed, all formal probability follows as a logical consequence of these axioms. Some of the most important simple results are summarised here. Throughout this section, assume that Ω is our collection of possible outcomes, A is a σ-alaebra over Ω and $PT.7$ is an associated nobabilitu distribution.

Many of these results simply demonstrate that things which we would intuitively want to be true of probabilities do, indeed, arise as logical consequences of this simple axiomatic framework.

Proposition 1.4: $P[\phi] = 0$.

Proposition 1.5: If $A_1, A_2, \ldots, A_n$ are pairwise disjoint elements of A, corresponding to mutually exclusive outcomes in our experiment, then

$$P[A_1 \cup A_2 \cup \ldots \cup A_n] = \sum_{i=1}^{n} P[A_i]$$

Proposition 1.6: If $A \in A$, then

$$P[A] = 1 - P[A]$$

Proposition 1.7: For any two events $A, B \in A$

$$P[A \cup B] = P[A] + P[B] - P[A \cap B]$$

Proposition 1.8: If $A, B \in A$ and $A \subset B$, then

$$P[A] \leq P[B]$$

Proposition 1.9 (Boole's Inequality): If $A_1, \ldots, A_n \in A$, then $P[A_1 \cup A_2 \cup \ldots \cup A_n] \leq P[A_1] + P[A_2] + \cdots + P[A_n]$.

Definition 1.2: (Probability Space): A probability space is the triple $(\Omega, A, P[].), where\Omega$ is a sample space, A is a σ-algebra over Ω, and $P[\ _0]$ is a probability function with domain A.

1.3 Conditional Probability and Independence

Sometimes it's possible to observe that one event has occurred. In this situation, we wish to have a model for the behaviour of the probability that other events compatible with B. Conditional probability is the appropriate language.

Definition 1.3: (Conditional Probability): Let A and B be events in A of the given probability space $(\Omega, A, P[.]). Theconditionalprobability$ of event A given event B, denoted by $P[A \mid B]$, is defined as

$P[A \mid B] = \frac{P[A \cap B]}{P[B]}$ if $P[B] > 0$,

and is left undefined when $P[B] = 0$.

Exercise1.3.1: Consider the experiment of tossing two coins, $\Omega = \{(H,H),(H,T),(T,H),(T,T)\}$, and assume that each point is equally likely. Find

(i.) The probability of two heads given a head on the first coin.

(ii.) The probability of two heads given at least one head.

Theorem 1.1: (Law of Total Probability): For a given probability space $(\Omega, A, P[].), if B_1, \ldots, B_n$ is a collection of mutually disjoint events in A satisfying

$$\Omega = \cup_{i=1}^{n} B_i,$$

i.e. $B_1, \ldots, B_n$ partition Ω and $P\left[B_j\right] > 0, j = 1, \ldots, n$, then for every $A \in A$,

$$P[A] = \sum_{j=1}^{n} P\left[A \cap B_j\right].$$

Conditional probability has a number of useful properties. The following elementary result is surprisingly important and has some far-reaching consequences.

Theorem 1.2 (Bayes' Formula): For a given probability space $(\Omega, A, P[\cdot])$, if $A, B \in A$ are such that $P[A] > 0$, $P[B] > 0$, then:

$$P[A \mid B] = \frac{P[B \mid A] P[A]}{P[B]}$$

Theorem 1.3 (Partition Formula): If $B_1, \ldots, B_n \in A$ partition Ω, then for any $A \in A$:

$$P(A) = \sum_{i=1}^{n} P(A \mid B_i) P(B_i)$$

Theorem 1.4 (Multiplication Rule): For a given probability space $(\Omega, A, P[\cdot])$, let $A_1, \ldots, A_n$ be events belonging to A for which $P[A_1, \ldots, A_{n-1}] > 0$, then
$P[A_1, A_2, \ldots, A_n] = P[A_1]P[A_2 \mid A_1] \cdot P[A_3 \mid A_1, A_2] \cdots \cdot P[A_n \mid A_1, \ldots, A_{n-1}]$.

Definition 1.4 (Independent Events): For a given probability space ($\Omega, A, P[\cdot]$), let A and B be two events in A. Events A and B are defined to be independent iff one of the following conditions is satisfied .

(i) $P[A \cap B] = P[A]P[B]$
(ii) $P[A \mid B] = P[A]$ if $P[B] > 0$
(iii) $P[B \mid A] = P[B]$ if $P[A] > 0$.

Exercise 1.3.2: Consider the experiment of rolling two dice. Let $A = \{$ total is odd $\}, B = \{6$ on the first die $\}, C = \{$ total is seven $\}$.

(i) Are A and B independent?
(ii) Are A and C independent?
(iii) Are B and C independent?

Definition 1.5 (Independence of Several Events): For a given probability space $(\Omega, A, P[\cdot])$, let $A_1, \ldots, A_n$ be events in A. Events $A_1, \ldots, A_n$ are defined to be independent iff

(i) $P[A_i \cap A_j] = P[A_i]P[A_j], i \neq j$
(ii) $P[A_1 \cap A_j \cap A_k] = P[A_1]P[A_j]P[A_k], i \neq j, j \neq k, i \neq k$
$\vdots$
n
$P[\cap_{i=1} A_i] = \prod_{i=1}^{n} P[A_i]$.

Random Variables

2.1 Random Variables and Cumulative Distribution Functions

We considered random events in the previous chapter: experimental outcomes which either do or do not occur. In general we cannot predict whether or not a random event will or will not occur before we observe the outcome of the associated experiment - although if we know enough about the experiment we may be able to make good probabilistic predictions. The natural generalization of a random event is a random variable: an object which can take values in the set of real numbers (rather than simply happening or not happening) for which the precise value which it takes is not known before the experiment is observed.

The following definition may seem a little surprising if you've seen probability only outside of measure-theoretic settings in the past. In particular, random variables are deterministic functions: neither random nor variable in themselves. This definition is rather convenient; all randomness stems from the underlying probability space and it is clear that random variables and random events are closely related. This definition also makes it straightforward to define multiple random variables related to a single experiment and to investigate and model the relationships between them.

Definition 2.1 (Random Variable): Given a probability space ($\Omega, A, P[.]$), a random variable, X, is a function with domain Ω and co-domain R (the real line) (i.e., $X: \Omega \rightarrow R$).

Example 2.1: Roll 2 dice $\Omega = \{(i, j); i, j = 1, \ldots, 6\}$. Several random variables can be defined, for example $X(i, j) = i + j$, also $Y((i, j)) = |i - j|$. Both, X and Y are random variables. X can take values $2, 3, \ldots, 12$ and Y can take values $0, 1, \ldots, 5$.

Definition 2.2 (Distribution Function): The distribution function of a random variable X, denoted by $F_X(\cdot)$ is defined to be the function $F_X: R \rightarrow [0,1]$ which assigns

$F_X(x) = P[X \leq x] = P[\{\omega : X(\omega) \leq x\}]$

for every $x \in R$.

Properties of $F_X(\cdot)$

(i) $\lim_{x \to -\infty} F_X(x) = 0$ and $\lim_{x \to +\infty} F_X(x) = 1$

(ii) $F_X(a) \leq F_X(b)$ for $a < b$ (monotonic and non-decreasing)

(iii) $F_X(\cdot)$ is continuous from the right

$\lim_{h \downarrow 0} F_X(x + h) = F_X(x)$

Any function $F(\cdot)$ with $F : R \to [0,1]$ satisfying the above properties is a distribution function for some random variable.

2.2 Density Functions

For two distinct classes of random variables, the distribution of values can be described more simply by using density functions. These classes are termed 'discrete' and 'continuous'.

Definition 2.3 (Discrete Random Variable): A random variable X is discrete if the range of X is countable (i.e. it is finite or isomorphic to the natural numbers).

Definition 2.4 (Discrete Density Function): If X is a discrete random variable with distinct values $x_1, x_2, \ldots, x_n$, ..., then the discrete density function of X is defined by

$$f_X(x) = \{P\left[X = x_j\right] \ if \ x = x_j, j = 1,2, \ldots, n, \ldots \ 0 \ \ if \ x \neq x_j$$

Exercise 2.2.1: Consider the experiment of tossing two dice. Let $X = \{$ total of upturned faces $\}$ and $Y = \{$ absolute difference of upturned faces$\}$.

(i) Give the probability function f_X and sketch it.

(ii) Give f_Y.

Definition 2.5: Any function $f : R \to [0,1]$ is defined to be a discrete density function if for some countable set $x_1, x_2, \ldots, x_n$, ...,

(i) $f\left(x_j\right) \geq 0, j = 1,2, \ldots$

(ii) $f(x) = 0$ for $x \neq x_j; j = 1,2, \ldots$

(iii) $\sum_j \ f\ \left(x_j\right) = 1$ where summation is over $x_1, x_2, \ldots, x_n$, ...

Definition 2.6 (Continuous Random Variable): A random variable X is called continuous if there exists a function $F_X(\cdot)$ such that

$$F_X(x) = \int_{-\infty}^{x} f_X\ (u) du \ for \ every \ x \in R$$

Definition 2.7 (Probability Density Function of a Continuous Random Variable): If X is a continuous random variable, the function f_X in $F_X(x) = \int_{-\infty}^{x} f_X\ (u) du$ is called the probability density function of X.

Definition 2.8: A function $f : R \to [0, \infty)$ is a probability density function if

(i) $f(x) \geq 0 \ \forall x$

(ii) $\int_{-\infty}^{+\infty} f(x) dx = 1$.

2.3 Expectations and Moments

Definition 2.9 (Expectation, Mean): Let X be a random variable. The mean of X, denoted by μ_X or $E[X]$ is defined by

(i) $E[X] = \sum_j \ x_j f_X\ \left(x_j\right)$ if X is discrete with mass points $x_1, x_2, \ldots, x_j$, ...

(ii) $E[X] = \int_{-\infty}^{\infty} x f_X(x)\ dx$ if X is continuous with density $f_X(x)$.

Intuitively, $E[X]$ is the centre of gravity of the unit mass that is specified bu the density function.

Exercise 2.3.1: Consider the experiment of rolling two dice. Let X denote the total of two dice and Y their absolute difference. Compute $E[X]$ and $E[Y]$.

Exercise 2.3.2: Let X be a continuous random variable with density

$$f_x(x) = \{\lambda e^{-\lambda x}\ \ if\ 0 \le x < \infty\ 0\ \ otherwise.$$

Compute $E[x]$ and $F_x(x)$.

Definition 2.10 (Variance): Let X be a random variable, and let $\mu_x = E[X]$. The variance of X, denoted by σ_x^2 or $Var[X]$ is defined by

(i) $Var[X] = \sum_j \left(x_j - \mu_X\right)^2 f_X\left(x_j\right)$ if X is discrete with mass points $x_1, x_2, \ldots, x_j, \ldots$

(ii) $Var[x] = \int_{-\infty}^{\infty} (x - \mu_x)^2 f_x(x) dx$ for continuous X with density $f_x(x)$.

Variance is a measure of spread or dispersion. If the values of a random variable X tend to be far from their mean, the variance of X will be larger than the variance of a comparable random variable Y whose values are typically nearer to the mean.

Definition 2.11 (Standard Deviation) : If X is a random variable, the standard deviation of X, denoted by σ_x, is defined as $\sigma_x = \sqrt{Var[x]}$.

2.4 Expectation of a Function of a Random Variable

Definition 2.12 (Expectation of a Function of a Random Variable): Let X be a random variable and $g(\cdot)$ a function $g: R \to R$. The expectation or expected value of the function $g(\cdot)$ of the random variable x, denoted by $E[g(x)]$ is defined by

(i) $E[g(x)] = \sum_j\ g\left(x_j\right) f_x\left(x_j\right)$ if x is discrete with mass points $x_1, x_2, \ldots, x_j, \ldots$, and provided the series is absolutely convergent.

(ii) $E[g(x)] = \int_{-\infty}^{\infty} g\ (x) f_x(x) dx$, for continuous x with density $f_x(x)$, provided the integral exists.

If $g(x) = x$, then $E[g(x)] = E[x] = \mu_x$. If $g(x) = (x - \mu_x)^2$, then $E[g(x)] = \sigma_x^2$.

Proposition 2.1 (Properties of Expectations): Expectations have a number of useful properties which aro very often useful:

(i) $E[c] = c$ for a constant c

(ii) $E[cg(x)] = cE[g(x)]$ for a constant c

(iii) $E[c_1 g_1(x) + c_2 g_2(x)] = c_1 E[g_1(x)] + c_2 E[g_2(x)]$

(iv) $E[g_1(x)] \le E[g_2(x)]$ if $g_1(x) \le g_2(x)\ \forall x$

Proposition 2.2 (Variance in Terms of Expectations) : If X is a random variable, then

$$Var[x] = E[(x - E[x])^2] = E[x^2] - (E[x])^2.$$

2.5 Two Important Inequality Results

Theorem 2.1 (Chebyshev's Inequality): Let X be a random variable and $g(\cdot)$ a non-negative function, then

$$P[g(x) \ge k] \le \frac{E[g(x)]}{k}\ \forall k > 0.$$

Corollary 2.1: If X is a random variable with finite variance, σ_X^2, then:

$$P[|x - \mu_x| \ge r\sigma_x] = P[(x - \mu_x)^2 \ge r^2 \sigma_x^2] \le \frac{1}{r^2}.$$

Note, that the last statement can also be written as

$$P[|x - \mu_x| \le r\sigma_x] \ge 1 - \frac{1}{r^2}.$$

Thus the probability that X falls within $r\sigma$ units of μ_X is greater than or equal to $1 - \frac{1}{r^2}$. For $r = 2$ one gets

$$P[\mu_x - 2\sigma_x < x < \mu_x + 2\sigma_x] \geq \frac{3}{4}.$$

or, for each random variable X having a finite variance, at least 3/4 of the mass of X falls within two standard deviations of its mean. Chebyshev's inequality gives a bound which does not depend on the distribution of X.

Theorem 2.2 (Jensen's Inequality): For a RV X with mean $E[X]$ and $g(\cdot)$ a convex continuous function,

$$E[g(X)] \geq g(E[X]).$$

Note, a convex function, g, is any function which lies beneath any of its chords. That is, given points a and b, with $a < b$, for any $\lambda \in [0,1]$:

$$g(\lambda a + (1-\lambda)b) \leq g((1-\lambda)a + \lambda b) \leq (1-\lambda)g(a) + \lambda g(b).$$

2.6 Moments and Moment Generating Functions

Definition 2.13 (Moments and Central Moments): For a RV X, the r^{th} moment of X is given by

$$\mu_r' = E[X^r]$$

If the expectation exists. For a RV X, the r^{th} central moment of X (the r^{th} moment about μ_1') is given by

$$\mu_r = E[(X-\mu_1')^r] = E\left[\sum_{i=0}^{r}\binom{r}{i}(\mu_1')^{r-i}X^i\right] = \sum_{i=0}^{r}\binom{r}{i}(\mu_1')^{r-i}E[X^i].$$

The First Four Central Moments.

Zero $\mu_1 = E[(X-\mu_X)] = 0$

Variance $\mu_2 = E[(X-\mu_X)^2]$

Skewness $\gamma_1 = \frac{\mu_3}{\sigma_x^3} = \frac{E[(X-\mu_X)^3]}{\sigma_x^3}$

$$"\,ET/V\,..\,147$$

Note that all odd central moments of X about μ_x are 0 if the density function is symmetrical about μ_x. The skewness is used to indicate whether a density is skewed to the left (value for skewness is negative) or skewed to the right (value for skewness is positive). The kurtosis is sometimes used to indicate that a density is more peaked (and heavier tailed) around its centre than the normal density.

The moment generating function gives us a function from which all the moments can be reconstructed via differentiation.

Definition 2.14 (Moment Generating Function (MGF)):

Let X be a R.V. with density $f_X(\cdot)$. We may define it's moment generating function, $m(t)$, as:

Discrete

$$m(t) = E[e^{tx}] = \sum_x e^{tx} f_X(x)$$

Continuous

$$m(t) = E[e^{tx}] = \int_{-\infty}^{\infty} e^{tx} f_X(x)dx$$

The MGF has the property that

$$\frac{d^r m(t)}{dt^r}\bigg|_{t=0} = \mu_r'$$

Theorem 2.3 (Equality of Distributions): Let X and Y be two random variables with densities $f_X(\cdot)$ and $f_Y(\cdot)$, respectively. Suppose that $m_X(t)$ and $m_Y(t)$ both exist and are equal for all t.

Then the two distribution functions $F_X(\cdot)$ and $F_Y(\cdot)$ are equal.

Exercise 2.6.1: Find the MGF of the binomial distribution

$$P(X=x) = \binom{n}{x}p^x(1-p)^{n-x}, x = 0,1,2,\dots,n$$

2.7 Other Distribution Summaries

Definition 2.15 (Quantile): For a R.V. X, the q^{th} quantile of η_q is the smallest number η satisfying

$$F_X(\eta) \geq q$$

Certain quantiles are often used in statistics and are given particular names:

Median $\eta_{0.5}$, a measure of the location of the distribution's centre.

Lower Quartile $\eta_{0.25}$ and Upper Quartile $\eta_{0.75}$ can be combined to give a measure of the spread of the distribution termed the Inter-quartile Range $\eta_{0.75} - \eta_{0.25}$ and together with the median provide some idea of the skewness of the distribution.

Another measure of the centre of a distribution is its mode: the point at which $f_X(\cdot)$ obtains its maximum.

Special Univariate Distributions

3.1 Discrete Distributions

3.1.1 Degenerate Distribution

(Degenerate at k) i.e.

$$P(X = k) = 1 \begin{cases} X \in \{k\} \\ \text{or } X(\Omega) = k \end{cases}$$

Then $F(x) = \begin{cases} 0 & \text{if } x < k \\ 1 & \text{if } x \geq k \end{cases}$

$E(X) = (k) \cdot 1 = k, Var(X) = E(X^2) - (E(X))^2$

$E(X^2) = k^2 \cdot 1 = k^2 \,, = k^2 - k^2 = 0$

$Var(X) = 0$. This property characterizes degenerate R.V.

3.1.2 Two Point Distribution

$X \in \{x_1, x_2\}$ say $x_1 < x_2$

Such that

$$P(X = x_1) = p, P(X = x_2) = 1 - p \; 0 < p < 1$$

Then

$$F(x) = \{0 \;\; if \; x < x_1 \; p \;\; if \; x_1 \leq x < x_2 \; 1 \;\; if \; x \geq x_2$$

$$E(X) = px_1 + (1 - p)x_2$$

$$M(t) = pe^{tx_1} + (1 - p)e^{tx_2}$$

$$Var(X) = p(1 - p)(x_1 - x_2)^2$$

Special Case:

If $x_1 = 1, x_2 = 0$

Then X is said to be Bernoulli R.V.

$$E(X) = p, Var(X) = p(1 - p)$$

Here, we can call it as probability of success is p and failure is $1 - p$

3.1.3 Uniform Distribution on n-Points

X is said to have a uniform distribution on n-points $\{x_1, \dots, x_n\}$

$$P(X = x_i) = \frac{1}{n} \; \forall i = 1,2, \dots, n$$

Then

$$E(X) = \frac{\sum x_i}{n} \quad M(t) = \sum_{i=1}^{n_{m_i}} \frac{e^{tx_i}}{n}$$

Special Case

If $x_i = i$

i.e. $X \in \{1,2, \dots, n\}$

Then $E(X) = \frac{n+1}{2}, Var(X) = \frac{n^2-1}{12}$

3.1.4 Binomial Distribution

X is said to have binomial with parameter p if its PMF is

$$P(X = k) = {}^{n}C_k p^k (1-p)^{n-k};\ k = 0,1, \dots, n\ 0 \quad \leq p \leq 1$$

We write

$$x \sim b(n, p)$$

$$E(X) = np \quad V(X) = npq \quad M(t) = (q + pe^t)^n \forall t$$

Note: Binomial distribution can also be considered as the distribution of sum of n independent & identical Binomial r.v. i.e. $b(1, p)$

Results:

(i) Let $X_i (i = 1,2, \dots, k)$ be independent R.V's with $X_i \sim b(n_i, P)$ then $X_1 + X_2 + \dots + X_k \sim b(n_1 + \dots + n_k, P)$ distribution.

In particular, take $n_i = n \ \forall i$.

(ii) Let X, Y be two independent, non-negative finite integer valued R.V.'s & let $Z = X + Y$. Then Z is a binomial R.V. with parameter p iff X & Y are binomial R.V. with same parameter p

(iii) If $X \sim b(n, p)$ then $n - X \sim b(n, q)$

3.1.5 Negative Binomial Distribution (Pascal or Waiting time distribution)

Let X denotes no. of failures that precede the r^{th} success

Then $X + r$ is total no. of experiments replication to get r successes.

$$\therefore \ P(x + x) = \left(\frac{x + r - 1}{x}\right) p_x^r (1-p)^x$$

We write,

$$X \sim NB(r, p)$$

$$E(X) = \frac{rq}{p}, M(t) = p^r (1 - qe^t)^{-r}$$

$$Var(X) = \frac{rq}{p^2} \ provided \ qe^t < 1$$

Note: If we are interested in no. of experiment trials needed.

Then $Y = X + r$

$\therefore P(Y = y) = \binom{y-1}{r-1} p^r (1-p)^{y-r}; y = r, r+1, \dots.$

Special Case:

If $r = 1$

then we say X is a geometric R.V.

Results:

(i) Let $X_1, \dots, X_k$ be independent $NB(r_i, p)$

then $\sum_{i=1}^{k} X_i$ follows $NB(r_1 + \dots + r_k, p)$

Corollary: Take $r_1 = 1$ then sum of independent geometric is binomial.

(ii) Let X and Y be independent $NB(r_1, p)$ and $NB(r_2, p)$. Then condition PMF of X given $X+Y=t$ is

$$P(X=x \mid X+Y=t)=\frac{\binom{x+r_1-1}{x}\binom{t+r_2-x-1}{t-x}}{\binom{t+r_1+r_2-1}{t}}$$

In particular, if $r_1=r_2=1$, then this distribution is uniform on t points.

(iii) If X has a geometric distribution then for any $m, n \in N \cup \{0\}$ (Memory Less) $P\{X>m+n \mid X>m\}=P\{X \geq n\}$

Converse is also true.

(iv) Let X be a non-negative integer valued R.V. satisfying

$$P\{X>m+1 \mid X>m\}=P\{X \geq 1\} \; \forall m$$

Then X must have geometric distribution

(v) Let $X_i (i=1, \ldots, n)$ be independent geometric R.V. with parameter p_i then $X_{(1)}=min\{X_i\}$ is also geometric with parameter

$$p=1-\prod_{i=1}^{n}(1-p_i)$$

3.1.6 Hyper-Geometric Distribution

A box contains N marbles of which M are marked. Now, n marbles are drawn. Let X denote no. of marked marbles drawn.

$$P(X=x)=\frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}; x \leq min(M, n) \;\; E(X)=\frac{n}{N}M \;\; Var(X)$$

$$=\frac{nM}{N^2(N-1)}(N-M)(N-n)$$

Results:

(i) Let X and Y be independent R.V.'s with distributions $b(m, p) \& b(n, p)$

Then conditional distribution of X given $X+Y$ is hyper geometric.

$$P(X=x \mid X+Y=N)=\frac{P(X=x, X+Y=N)}{P(X+Y=N)}$$

$$=\frac{P(X=x)P(Y=N-x)}{P(X+Y=N)}$$

$$=\frac{{}^{m}C_x p^x q^{m-x} \, {}^{n}C_{N-x} p^{N-x} q^{n-N+x}}{{}^{m+n}C_N p^N q^{m+n-N}}$$

$$=\frac{{}^{m}C_x \, {}^{n}C_{N-x}}{{}^{m+n}C_N}$$

3.1.7 Poisson Distribution

A R.V. is said to be a Poisson R.V. with parameter $\lambda>0$ if its PMF is given by

$P(X=k)=\frac{e^{-\lambda}\lambda^k}{k!}, k=0,1,2, \ldots.$

$$E(X)=\lambda, M(t)=e^{\lambda(e^t-1)}$$
$$Var(X)=\lambda$$

Results:

(i) Let $X_1, X_2, \ldots, X_n$ be independent Poisson R.V.'s $X_k \sim P(\lambda_k), k=1,2, \ldots, n$

Then

$$S_n = X_1 + \cdots + X_n \ is\ a\ P(\lambda_1 + \cdots + \lambda_n)\ R.V.'s$$

Converse is also true.

(ii) Let $X \& Y$ be independent R.V.'s with $P(\lambda_1)$ and $P(\lambda_2)$ respectively then conditional distribution of X given $X + Y$ is binomial.

(Converse is true)

Uses of the Poisson Distribution

For large n, and small $p, X \sim Bin(n,p)$ is approximately distributed as Poi (np). This is sometimes termed the "law of small numbers".

A Poisson Process with rate λ per unit time is such that

(i) X, the number of occurrences of an event in any given time interval of length t is $Poi(\lambda t)$.

(ii) The number of events in non-overlapping time intervals are independent random variables (see later).

3.1.8 Multi-Nomial Distribution (Generalized Binomial Distribution)

Let $x_1, x_2, \ldots, x_{k-1}$ be non-negative integers such that $x_1 + x_2 + \cdots + x_{k-1} \leq n$ then probability that exactly x_i trials terminate in $A_i (i = 1,2,\ldots,k-1)$ & hence that $x_k = n - (x_1 + x_2 + \cdots + x_{k-1})$ trials is

$$\frac{n!}{x_1!\,x_2! \ldots x_k!} p_1^{x_1} p_2^{x_2} \ldots p_k^{x_k}$$

$$\sum_{i=1}^{k} x_i = n$$

$$M(t) = (p_1 e^t + p_2 e^t + \cdots + p_{k-1} e^t + p_k)^n$$

$$E(e^{t_1 x_1 + t_2 x_2 + \cdots + t_k x_k}) \forall t_1, t_2, \ldots, t_k \in R$$

$$E(X_i) = np_i, Var(X_i) = np_i(1 - p_i)$$

$$Cov(X_i, X_k) = -np_i p_k$$

Summary

S.No.	Distribution	PMF	$E(X)$	Var(X)	$M(t)$
1	Poisson $X \sim P(\lambda)$	$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$ $k = 0,1,2,\ldots$	λ	λ	$e^{\lambda(e^t - 1)}$
2	Binomial $X \sim B(n,p)$	$P(X = k) = {}^nC_k p^k q^{n-k}$ $k = 0,1,2,\ldots,n$	np	npq	$(q + pe^t)^n$
3	Uniform $X \sim U(1,n)$	$P(X = k) = \frac{1}{n}$ $k = 1,2,\ldots,n$	$\frac{n+1}{2}$	$\frac{n^2 - 1}{12}$	$\frac{1}{n}\sum_{k=1}^{n} e^{tk}$
4	Two Point	$P(X = a) = p$ $P(X = b) = 1 - p$	$ap + b(1 - p)$	$p(1 - p)(a - b)^2$	$pe^{at} + (1 - p)e^{bt}$
5	Negative Binomial $X \sim NB(r,p)$	$P(X = x) = \binom{x + r - 1}{x} p^r (1 - p)^x$	$\frac{rq}{p}$	$\frac{rq}{p^2}$	$p^r(1 - qe^t)^{-r}$; $qe^t < 1$

6	Geometric $X \sim G(1,p)$	$P(X=x)=pq^{x-1}$	$\frac{q}{p}$	$\frac{q}{p^2}$	$\frac{p}{1-qe^t}$; $qe^t<1$
7	Hypergeometric	$P(X=x)$ $=\frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$	$\frac{n}{N}M$	$\frac{nM}{N^2(N-1)}(N-M)(N-n)$	

3.2 Continuos Distributions

3.2.1. Uniform Distribution

X is said to have uniform distribution on $[a,b]$ if its PDF is given by

$$f(x)=\begin{cases}\frac{1}{b-a} & ;a\le x\le b\\ 0 & ;\ \text{otherwise}\end{cases}$$

$E(X)=\frac{a+b}{2}, Var(X)=\frac{(b-a)^2}{12}$

$M(t)=\frac{1}{t(b-a)}(e^{tb}-e^{ta}); t\neq 0$

Results: Let X be an R.V. with a continuous D.F. F, then F(X) has the uniform distribution on [0,1]

3.2.2. Gamma Distribution

An R.V. X is said to have gamma distribution with parameters α and β if its PDF is

$$f(x)=\begin{cases}\frac{1}{(\Gamma\alpha)\beta^\alpha}x^{\alpha-1}e^{-x/\beta}, & 0<x\\ 0 \quad ; & \text{otherwise}\end{cases}$$

We write $X\sim G(\alpha,\beta)$

$$E(X)=\alpha\beta\ Var(X)=\alpha\beta^2$$

$$M(t)=\frac{1}{(1-\beta t)^\alpha}, t<\frac{1}{\beta}$$

Special Case:

(a) When $\alpha=1$, then we say X follows exponential distribution with parameter β

$$f(x)=\begin{cases}\frac{1}{\beta}e^{-x/\beta} & ;x>0\\ 0 & ;x<0\end{cases}$$

or $\lambda=\frac{1}{\beta}$

$$f(x)=\begin{cases}\lambda e^{-\lambda x} & ;\quad x>0\\ 0 & ;\ \text{otherwise}\end{cases}$$

$$E(X)\ =\beta=\frac{1}{\lambda}$$

(b) When $\alpha=\frac{n}{2}$ ($n>0$ integer) & $\beta=2$

Then

$$f(x)=\begin{cases}\frac{1}{\Gamma_{n/2}2^{n/2}}e^{-x/2}x^{n/2-1}; & 0<x<\infty\\ 0; & \text{otherwise}\end{cases}$$

is said chi-square $\chi^2(n)$ distribution.

$$E(X) = n, Var(X) = 2n$$

$$M(t) = \frac{1}{(1-2t)^{n/2}}, t < \frac{1}{2}$$

Results: Let $X_i (i = 1, \dots, n)$ be independent R.V. such that $X_i \sim G(\alpha_i, \beta)$, then

$$S_n = \sum_{i=1}^{n} X_i \sim G\left(\sum_{i=1}^{n} \alpha_i, \beta\right) R.V.$$

Corollary:

(i) Take $\alpha_i = 1 \; \forall i$
Then $S_n \sim G(n, \beta)$
i.e. sum of exponential is Gamma

(ii) Let $X \sim G(\alpha_1, \beta) \& Y \sim G(\alpha_2, \beta)$ be independent R.V. then
$X + Y$ and $\frac{X}{Y}$ are independent
or $X + Y$ and $\frac{X}{X+Y}$ all independent
conversely also true.

(iii) Memory less property of exponential
$P(X > r + s \mid X > s) = P(X > r)$
where $X \sim exp\ (\lambda)$

(iv) If X & Y are independent exponential R.V.'s with parameter β, then
$Z = \frac{X}{X+Y}$ has a $U(0,1)$ distribution

3.2.3. Beta Distribution

An R.V. X is said to have beta distribution with parameters $\alpha \& \beta (\alpha > 0, \beta > 0)$ if its PDF is

$$f(x) = \begin{cases} \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}; & 0 < x < 1 \\ 0; & \text{otherwise} \end{cases}$$

We write $X \sim B(\alpha, \beta)$

$E(X) = \frac{\alpha}{\alpha+\beta}$ $\qquad$ $V(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$

Note: If $\alpha = \beta = 1$, we have $U(0,1)$

Results:

(i) $X \sim B(\alpha, \beta)$ then $1 - X \sim B(\beta, \alpha)$

(ii) Let $X \sim G(\alpha_1, \beta) \& Y \sim G(\alpha_2, \beta)$ be independent then

$$\frac{X}{X+Y} \sim B(\alpha_1, \alpha_2) \;\; R.V.$$

3.2.4. Normal Distribution (Gaussian Law)

(a) An R.V. X is said to have a standard normal distribution
if its PDF is

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, -\infty < x < \infty$$

we write $X \sim N(0,1)$

(b) An R.V. is said to have normal distribution with parameters μ & $\sigma(>0)$ if

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

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

$$M(t) = exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right)$$

Central moments

$$i.e. \quad E[(X-\mu)^n] = 0\ if\ n\ is\ odd \quad = [(2n-1)(2n-3)\dots 3.1]\sigma^{2n}\ if\ n\ is\ even$$

Results:

(i) Let $X_1, X_2, \dots, X_n$ be independent R.V.'s such that

$$X_k \sim N(\mu_k, \sigma_k^2), k = 1, \dots, n$$

then

$$S_n = \sum_{k=1}^{n} X_k\ is\ N\left(\sum_{k=1}^{n}\mu_k, \sum_{k=1}^{n}\sigma_k^2\right)$$

Corollary:

(a.) $X_i \sim N(\mu, \sigma^2)$

Then

$$S_n \sim N(n\mu, n\sigma^2)\ \ \& \frac{S_n}{n} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

(b.) If $X_i \sim N(0,1)(i = 1, \dots, n)$ are independent, then

$$\frac{S_n}{\sqrt{n}} \sim N(0,1)$$

(ii) Let $X \& Y$ be independent R.V.'s then $X + Y$ is normal if $X \& Y$ is normal

(iii) Let X & Y be independent R.V.'s with $N(0,1)$ then $X + Y$ and $X - Y$ are independent.

(iv) Let $X_1 \& X_2$ are independent $N(\mu_1, \sigma^2) \& N(\mu_2, \sigma^2)$ then $X_1 - X_2$ and $X_1 + X_2$ are independent.

(v) (a) $X \sim N(0,1) \Rightarrow X^2 \sim N^2(1)$

(b) $X \sim N(\mu, \sigma^2) \Rightarrow aX \sim N(a\mu, a^2\sigma^2)$

$$aX + b \sim N(a\mu + b, a^2\sigma^2)$$

(vi) $X \sim N(\mu, \sigma^2) \Rightarrow \frac{X-\mu}{\sigma} \sim N(0,1)$

(vii) $X \& Y$ be i.i.d. $N(0, \sigma^2)$ R.V.'s then

$\frac{X}{Y} \sim$ Cauchy (1,0)

$\frac{X}{|Y|} \sim$ Cauchy (1,0)

$\frac{|X|}{|Y|}$ has PDF $\frac{2}{\pi(1+z^2)}, 0 < z < \infty$

3.2.5. Cauchy Distribution

An R.V. is said to have Cauchy distribution with parameters μ and θ if its PDF is

$$f(x) = \frac{\mu}{\pi} \frac{1}{\mu^2 + (x-\theta)^2}; -\infty < x < \infty\ \mu > 0$$

We write $X \sim C(\mu, \theta)$

Results:

(i) Moments of order ≥ 1 doesn't exist.

(ii) $X \sim C(\mu_1, \theta_1) \& Y \sim C(\mu_2, \theta_2)$ be independent then

$$X + Y \sim C(\mu_1 + \mu_2, \theta_1 + \theta_2)$$

(iii) $X \sim C(1,0)$ iff $\frac{1}{X} \sim C(1,0)$

(iv) If $X \sim U\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$ R.V. then $Y = tan\ X$ is Cauchy R.V.

Joint And Conditional Distributions

4.1 Joint Distributions

Definition 4.1 (Joint Distribution Function): For random variables $X_1, \dots, X_k$ all defined on the same $(\Omega, A, P[\cdot])$, the function $F: R^k \to [0,1]$

$$F_{X_1,\dots,X_k}(x_1, \dots, x_k) = P[X_1 \le x_1; \dots, X_k \le x_k]\ \forall (x_1, \dots, x_k)$$

is called the joint distribution function.

Definition 4.2 (Marginal Distribution Function): For $F_{X_1,\dots,X_k}$ and $X_{i_1}, \dots, X_{i_m}$ a strict subset of $X_1, \dots, X_k$, the function $F_{X_{i_1},\dots,X_{i_m}}$ is called a marginal distribution function.

Definition 4.3 (Joint Discrete Density Function): For a k-dimensional discrete random variable $(X_1, \dots, X_k)$, the function

$$f_{X_1,\dots,X_k}(x_1, \dots, x_k) = P[X_1 = x_1; \dots.; X_k = x_k] \forall (x_1, \dots, x_k)$$

is called the joint discrete density function, joint density function, or joint probability function.

Definition 4.4 (Marginal Discrete Density Function): For the joint discrete density function $f_{X_1,\dots,X_k}$ and $X_{i_1}, \dots, X_{i_m}$ a strict subset of $X_1, \dots, X_k$, the function $f_{X_{i_1},\dots,x_{i_m}}$ is called a marginal discrete density function.

Definition 4.5 (Joint Continuous Density Function): For a k-dimensional random variable $(X_1, \dots, X_k)$, the function $f_{X_1,\dots,X_k}(x_1, \dots, x_k) \ge 0$ such that

$$f_{X_1,\dots,X_k}(x_1, \dots, x_k) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \dots \int_{-\infty}^{x_k} f_{X_1,\dots,X_k}(u_1, \dots, u_k) du_1 \dots du_k$$

for all $(x_1, \dots, x_k)$ is called a joint probability density function.

Definition 4.6 (Marginal Density Function): For the joint continuous density function $f_{X_1,\dots,X_k}$ and $X_{i_1}, \dots, X_{i_m}$ a strict subset of $X_1, \dots, X_k$, the function $f_{X_{i_1},\dots,X_{i_m}}$ is called a marginal density function.

4.2 Special Multivariate Distributions

There are many ways to construct multivariate distributions and a great many named multivariate distributions exist. This section provides a brief overview.

summary of a small number of the most ubiquitous multivariate distributions.

4.2.1. Multinomial Distribution

The multivariate distribution is a generalisation of the binomial distribution.

$$f_{X_1,\dots,x_k}(x_1, \dots, x_k) = Mult(x; n, p_1, \dots, p_{k+1}) = \frac{n!}{\prod_{i=1}^{k+1} x_i!} \prod_{i=1}^{k+1} p_i^{x_i}$$

Where $x_i = 0, \dots, n$ and $\sum_{i=1}^{k+1} x_i = n$. (n is fixed, so the value of x_{k+1} is determined by values of $x_1, \dots, x_k$).

Note that the i^{th} element of X, X_i, has a Bin (n, p_i) marginal distribution.

4.2.2. Bi-Variate Normal Distribution

The normal distribution can be straightforwardly extended to a bivariate random variable by considering to random variables which are correlated and which each is marginally normal:

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} exp\,\{-\frac{1}{2(1-\rho)}[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 +$$

$$\left(\frac{x_2-\mu_2}{\sigma_2}\right)^2 - 2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right)]\}$$

for $-\infty < x_1, x_2, \mu_1, \mu_2 < \infty, \sigma_1, \sigma_2 > 0, -1 < \rho < 1$.

(a) ρ is the correlation coefficient
(b) for $\rho = 0$ the bi-variate normal density is the product of two uni-variate normal densities; this corresponds to X_1 and X_2 being independent Normal random variables.
(c) $x + y$ and $x - y$ are independent iff $\sigma_1^2 = \sigma_2^2$
(d) x and y are independent iff $\rho = 0$

Extension to Multivariate Normal Distribution: In fact, it's straightforward to extend the normal distribution to vectors or arbitrary length, the multivariate normal distribution has density.

$$f(x; \mu, \Sigma) = N(x; \mu, \Sigma) = \left(\frac{1}{\sqrt{2\pi}}\right)^r |\Sigma|^{-1/2} exp\left\{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right\}$$

Where

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_r \end{pmatrix}, \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_r \end{pmatrix}, \Sigma = E[(X-\mu)(X-\mu)^T]$$

Note that x is a vector; it has mean μ which is itself a vector and Σ is the variance-covariance matrix. If x is k-dimensional then Σ is a $k \times k$ matrix.

4.3 Conditional Distributions and Densities

Given several random variables how much information does knowing one provide about the others? The notion of conditional probability provides an explicit answer to this question.

Definition 4.7 (Conditional Discrete Density Function): For discrete random variables with x and Y with probability mass points $x_1, x_2, \dots, x_m$ and $y_1, y_2, \dots, y_n$,

$$f_{Y|X}(y_j \mid x_i) = \frac{P[X = x_i; Y = y_j]}{P[X = x_i]} = P[Y = y_j \mid X = x_i]$$

is called the conditional discrete density function of Y given $X = x$.

Definition 4.8 (Conditional Discrete Distribution): For jointly discrete random variables X and Y,

$$f_{Y|X}(y_j \mid x_i) = P[Y = y_j \mid X = x_i] = \textstyle\sum_{y_j \in S_Y} f_{Y|X}(y_j \mid x_i)$$

is called the conditional discrete distribution of Y given $X = x$.

Definition 4.9 (Conditional Probability Density Function): For continuous random variables X and Y with joint probability density function $f_{X,Y}(x, y)$,

$f_{Y|X}(y \mid x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$, if $f_X(x) > 0$

Where $f_X(x)$ is the marginal density of X.

Conditional Distribution For jointly continuous random variables X and Y, $f_{Y|X}(y \mid x) = \int_{-\infty}^{\infty} f_{Y|X}(z \mid x)dz\ \forall x$ such that $f_X(x) > 0$.

4.4 Conditional Expectation

We can also ask what the expected behaviour of one random variable is, given knowledge of the value of a second random variable and this gives rise to the idea of conditional expectation.

Definition 4.10 (Conditional Expectation): The conditional expectation in discrete and continuous cases corresponds to an expectation with respect to the appropriate conditional probability distribution:

Discrete

$$E[Y \mid X = x] = \sum_{all\, y} y f_{Y|X}(Y = y \mid X = x)$$

Continuous

$E[Y \mid X = x] = \int_{-\infty}^{\infty} y f_{Y|X}(Y \mid X = x)dy.$

Note that before x is known to take the value $x, E[Y \mid X]$ is itself a random variable being a function of the random variable X. We might be interested in the distribution of the random variable $E[Y \mid X]$.

Theorem 4.1 (Tower Property of Conditional Expectation): For any two random variables X_1 and X_2

$$E\big[E[X_1 \mid X_2]\big] = E[X_1]$$

Exercise 4.4.1: Suppose that $\theta \sim U[0,1]$ and $(X \mid \theta) \sim Bin(2, \theta)$. Find $E[X \mid \Theta]$ and hence or otherwise show that $E[X] = 1$.

4.5 Conditional Expectations of Functions of Random Variables

By extending the theorem on marginal expectations we can relate the conditional and marginal expectations of functions of random variables (in particular, their variances).

Theorem 4.2: (Marginal Expectation of a Transformed Random Variables): For any random variables X_1 and X_2, and for any function $h(\cdot)$,

$E\big[E[h(X_1) \mid X_2]\big] = E[h(X_1)].$

Theorem 4.3 (Marginal Variance): For any random variables X_1 and X_2,

$Var(X_1) = E[Var(X_1 \mid X_2)] + Var(E[X_1 \mid X_2]).$

4.6 Independence of Random Variables

Whilst the previous sections have been concerned with the information that one random variable carries about another, it would seem that there must be pairs of random variables which each provide no information whatsoever about the other. It is, for example, difficult to imagine that the value obtain when a die is rolled in Coventry will tell us much about the outcome of a coin toss taking place at the same time in Lancaster.

There are two equivalent statements of a property termed stochastic independence which capture precisely this idea. The following two definitions are equivalent for both discrete and continuous random variables.

Definition 4.11: (Stochastic Independence): Definition 1 Random variables $X_1, X_2, \ldots, X_n$ are stochastically independent iff

$$F_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = \prod_{i=1}^{n} F_{X_i}(x_i)$$

Definition 2: Random variables $X_1, X_2, \ldots, X_n$ are stochastically independent iff

$$f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = \prod_{i=1}^{n} f_{X_i}(x_i)$$

If X_1 and X_2 are independent then their conditional densities are equal to their marginal densities

4.7 Covariance and Correlation

Having established that sometimes one random variable does convey information about another and in other cases knowing the value of a random variable tells us nothing useful about another random variable it is useful to have mechanisms for characterising the relationship between pairs (or larger groups) of random variables.

Definition 4.12 (Covariance and Correlation): Covariance: For random variables X and Y defined on the same probability space

$$Cov[X,Y] = E[(X-\mu_X)(Y-\mu_Y)] \quad = E[XY] - \mu_X\mu_Y$$

Correlation: For random variables X and Y defined on the same probability space

$$\rho[X,Y] = \frac{Cov[X,Y]}{\sigma_X\sigma_Y} = \frac{Cov[X,Y]}{\sqrt{Var[X]}\sqrt{Var[Y]}}$$

provided that $\sigma_X > 0$ and $\sigma_Y > 0$.

Theorem 4.4 (Cauchy-Schwarz Inequality): Let X and Y have finite second moments. Then

$$(E[XY])^2 \leq E[X^2]E[Y^2]$$

with equality if and only if $P[Y = cX] = 1$ for some constant c.

4.8 Transformation of Random Variables: $Y = g(X)$

Theorem 4.5 (Distribution of a Function of a Random Variable): Let X be a random variable and $Y = g(X)$ where g is injective (i.e. it maps at most one x to any value y). Then

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

given that $(g^{-1}(y))'$ exists and $(g^{-1}(y))' > 0\ \forall y$ or $(g^{-1}(y))' < 0\ \forall y$. If g is not injective (one-to-one) there may be values of y for which there exists no x such that $y = g(x)$. Such points clearly have density zero.

When the conditions of this theorem are not satisfied it is necessary to be a little more careful. The most general approach for finding the density of a transformed random variable is to explicitly construct the distribution function of the transformed random variable and then to use the standard approach to turn the distribution function into a density (this approach is discussed in Larry Wasserstein's "All of Statistics").

Exercise 4.8.1: Let X be distributed exponentially with parameter α, that is

$$f_X(x) = \begin{cases} \alpha e^{-\alpha x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

Find the density function of

(i) $Y = g(X)$ with $g(X) = \begin{cases} 0 & \text{for } x < 0 \\ 1 - e^{-\alpha x} & \text{for } x \geq 0 \end{cases}$

(ii) $Y = X^\beta, \beta > 0$

(iii) $Y = g(X)$ with $g(X) = \begin{cases} 0 & \text{for } x < 0 \\ x & \text{for } 0 \leq x \leq 1 \\ 1 & \text{for } x > 1 \end{cases}$

Theorem 4.6 (Probability Integral Transformation): If X is a random variable with continuous $F_X(x)$, then $U = F_X(X)$ is uniformly distributed over the interval $(0,1)$.

Conversely if U is uniform over $(0,1)$, then $X = F_X^{-1}(U)$ has distribution function F_X.

4.9 Moment-Generating-Function Technique

The following technique is but one example of a situation in which the moment generating function proves invaluable.

Function of a Variable. For $Y = g(X)$ compute

$$m_Y(t) = E[e^{tY}] = E\left[e^{tg(X)}\right]$$

If the result is the MGF of a known distribution then it will follow that Y has that distribution.

Sums of Independent random variables. For $Y = \sum_{i=1}^{n} X_i$, where the X_i are independent random variables for which the MGF exists $\forall -h < t < h, h > 0$

$$m_Y(t) = E\left[e^{t\sum_{i=1}^{n} x_i}\right] = \prod_{i=1}^{n} m_{X_i}(t) \ for \ -h < t < h$$

Thus $\prod_{i=1}^{n} m_{X_i}(t)$ may be used to identify the distribution of Y as above.

Inference

5.1 Sample Statistics

Suppose we select a sample of size n from a population of size N. For each i in $\{1, \dots, n\}$, let X_i be a random variable denoting the outcome of the i^{th} observation of a variable of interest. For example, X_i might be the height of the i^{th} person sampled. Under the assumptions of simple random sampling, the X_i are independent and identically distributed (iid).

Therefore, if the distribution of a single unit sampled from the population can be characterized by a distribution with density function f, the marginal density function of each X_i is also f and their joint density function g is a simple product of their marginal densities:

$$g(x_1, x_2, \dots, x_n) = f(x_1)f(x_2)\cdots f(x_n)$$

In order to make inferences about a population parameter, we use sample data to form an estimate of the population parameter. We calculate our estimate using an estimator or sample statistic, which is a function of the X_i. We have already seen examples of sample statistics, for example the sample mean

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

where n is the size of the sample is an estimator of the population mean, e.g. for discrete X

$$\mu_X = \sum_{x_j} x_j P\left[X = x_j\right]$$

where N is the number of distinct values which it is possible for an X_i to take.

5.2 Sampling Distributions

Since an estimator $\hat{\theta}$ is a function of random variables, it follows that $\hat{\theta}$ is itself a random variable and possesses its own distribution. The probability distribution of an estimator itself is called a sampling distribution.

Proposition 5.1 (Distribution of The Sample Mean): Let $\bar{X}$ denote the sample mean of a random sample of size n from a normal distribution with mean μ and variance σ^2. Then

$$X^{-} \sim N\left(\mu, \frac{\sigma^2}{n}\right).$$

Theorem 5.1 (Central Limit Theorem):

Let f be a density function with mean μ and finite variance σ^2. Let $\bar{X}$ be the sample mean of a random sample of size n from f and let

$$Z_n = \frac{x^{-} - E[x^{-}]}{\sqrt{Var[x^{-}]}} = \frac{x^{-} - \mu}{\sigma/\sqrt{n}}.$$