On an Index of Discrimination

S. Subramanian*

Abstract

This note is concerned with deriving an exceedingly simple index of relative disadvantage, which seeks to measure the extent of group-related inequality in a population, when the latter is classified, after a binary fashion, according to identity characteristic and welfare status. Empirical examples are provided of the discrimination suffered by the Scheduled Castes and Tribes of India in access to various welfare outcomes.

Introduction

The phenomenon of discriminatory barriers to entry is one which is experienced, in a variety of social and economic situations, by those groups of individuals that are relatively 'disadvantaged' or 'underprivileged' in the socio-economic hierarchy of the reference society. Common examples would include discrimination in the job market as between males and females, discrimination in access to rented housing as between members of majority communities and members of minority communities, discrimination in access to occupational categories as between 'forward caste' and 'backward caste' persons, and so on.

The examples can be readily multiplied, but what they have in common is the following. Suppose a given set of individuals to be classified according to two broad attributes, A and B, where A is an *identity characteristic* (such as caste, religion, gender, etc.), and B is a *well-being characteristic* (such as literacy status, income status, housing quality status, etc.). Then, it is frequently the case that one observes a high order of concentration of certain 'identity-types' in certain 'well-being slots.' A quantitative measure of such an asymmetry in access would serve the purpose of a useful summary statistic of group-related discrimination.

The object of this note is to develop, in a preliminary way, one very elementary real-valued index of discrimination, which shall be designated by δ. The index δ, it should be pointed out, works for a two-way binary classification of the population, that is, for a situation in which categorisation of the population according

* Madras Institute of Development Studies, 79, Second Main Road, Gandhinagar, Adyar, Chennai - 600 020, Tamilnadu, India. E-mail: subbu@mids.tn.nic.in.

I am grateful to K. Nagaraj and D. Jayaraj for useful discussions on the subject of this note; to the latter for guidance on data and data sources; and to R. Dharumaperumal for assistance with the word processing of the paper.

to the identity attribute A precipitates two groups and, likewise, categorisation according to the well-being attribute B precipitates two groups. While this makes for a potentially restrictive outcome, it can nevertheless find useful application in a variety of real-life situations.

The Discrimination Index δ

Consider some identity attribute A (say, caste) which divides the population into two groups (say 'backward' and 'forward' castes respectively), and some well-being attribute B (say income status) which also divides the population into two groups (say 'low' and 'high' income categories respectively). Let t_{ij} be the proportion of the population belonging to identity category i (=1,2) and well-being category j (=1,2). The share of identity category i (=1,2) in the total population is designated by a_i, and that of well-being category j (=1, 2) by b_j. It is immediate that $t_{11} + t_{12} = a_1$; $t_{21} + t_{22} = a_2$; $t_{11} + t_{21} = b_1$; and $t_{12} + t_{22} = b_2$. Further, $a_1 + a_2 = b_1 + b_2 = 1$.

It seems reasonable to suggest that if membership to any given identity category has no influence on the prospects of membership to any given well-being category, then persons belonging to any given identity category should be distributed between the two well-being categories in the same proportion which the population of the relevant identity category bears to the total population. That is, one could envisage an 'ideal' or 'normative' situation of 'perfect symmetry of access,' wherein the t_{ij} are given by:

(1) $t^{\#}_{ij} = a_i b_j$, i = 1, 2; j = 1, 2.

Next, consider the quantity δ_{ij}, which is given by:

(2) $\delta_{ij} = \left| t_{ij} / t^{\#}_{ij} - 1 \right|$; i = 1, 2; j = 1, 2.

δ_{ij} is simply the proportionate deviation (in absolute terms) of t_{ij} from its 'normative' value $t^{\#}_{ij}$. If we were to sign δ_{ij}, then $\delta_{ij} > 0$ would imply that identity category i is 'over-represented' in well-being category j, and $\delta_{ij} < 0$ would imply that identity category i is 'under-represented' in well-being category j. The set of numbers $\{\delta_{11}, \delta_{12}, \delta_{21}, \delta_{22}\}$ thus presents a comprehensive picture of the extent of access which each identity category has to each well-being category.

To obtain a more aggregative picture, define the quantity $D_i = \sum_{j=1}^{2} \delta_{ij}$, i = 1, 2. D_i can be viewed as an *index of asymmetry of access for identity category i*, or simply as *i's index of discrimination*[1]. Notice that if $t_{ij} = t^{\#}_{ij}$ for all j, then it would be the case that $\delta_{ij} = 0$ for all j, and $D_i = 0$. Thus, the case of *perfect* symmetry of access for identity category i would correspond to $D_i = 0$ — which is, of course, the minimum value that D_i can assume. (It should be clear now why the δ_{ij} have been presented in absolute terms: if the δ_{ij} were to be signed, then a high positive value for one of the δ_{ij} and an equally high negative value for the other would, upon algebraic summation, yield a value of D_i which is zero; to interpret from this that identity category i enjoys symmetric access is seriously misleading).

We now have two 2-vectors: the vector **D** = (D_1, D_2), which is a vector of *actual* indices of discrimination for each identity category i = 1, 2; and the zero-vector **0** = (0,0), which is the vector that would correspond to *perfect* symmetry of access ('no discrimination') for each identity category i = 1, 2. To assess how far the actual outcome is from the ideal one, we would need to invoke some notion of *distance*. In this context, the notion of the *Euclidean norm* comes readily to mind. For any two vectors **x** = (x_1, x_2) and **y** = (y_1, y_2) belonging to two-dimensional Euclidean space, the distance between **x** and **y** is defined in terms of the Euclidean norm:

$$||\mathbf{x} - \mathbf{y}|| = \left[\sum_{k=1}^{2}(x_k - y_k)^2\right]^{1/2}.$$

Given the vectors **D** and **0**, let us define $\mathbf{D}^*$ to be that **D** which maximises ||**D - 0**||. Then, it is readily clear that we can write a simple, normalised, aggregate index of discrimination for the society as a whole as

(3) $\delta = ||\mathbf{D} - \mathbf{0}||/||\mathbf{D}^* - \mathbf{0}||$.

In (3), the aggregate index of discrimination δ is expressed as the ratio of the distance of **D** from **0** to the maximum possible distance of **D** from **0**. It remains to be specified how $\mathbf{D}^*$ is to be derived. To this end, consider the following.

First, let us designate ||**D - 0**|| by V. Then, in view of the earlier definitions of **D** and **0**, it is clear that

(4) $V = \left[\left\{\left| t_{11}/t_{11}^{\#} - 1\right| + \left| t_{12}/t_{12}^{\#} - 1\right|\right\}^2 + \left\{\left| t_{21}/t_{21}^{\#} - 1\right| + \left| t_{22}/t_{22}^{\#} - 1\right|\right\}^2\right].$

Making use of the fact that $\sum_{j=1}^{2} t_{ij} = a_i$ (i=1, 2); $\sum_{i=1}^{2} t_{ij} = b_j$ (j=1, 2); $a_1 + a_2 = b_1 + b_2 = 1$; and $t_{ij}^{\#} \equiv a_i b_j$ (i =1, 2; j =1, 2), it is possible, through routine manipulation, to show that (4) can be written equivalently as:

(5) $V = [|t_{11} - a_1 b_1]/a_1 a_2 b_1 b_{2.}$

Recalling that $\mathbf{D}^*$ is that **D** vector which maximises $||\mathbf{D} - \mathbf{0}|| \equiv V$, in order to obtain $\mathbf{D}^*$ we must, in view of (5), solve the following optimisation problem:

(6) Maximise $V = [|t_{11} - a_1 b_1|]/a_1 a_2 b_1 b_2$
$\{t_{11}\}$
subject to: $0 \le t_{11} \le \min\{a_1, b_1\}$.

[The upper bound constraint on t_{11} in problem (6) is dictated by the requirement that $t_{11} + t_{12} = a_1$ and $t_{11} + t_{21} = b_1$, which implies that t_{11} cannot exceed the smaller of a_1 and b_1.]

From inspection of problem (6), it is immediate that the optimal solution to the problem is furnished by

(7) $t_{11}^* = 0$ if $\min\{a_1, b_1\} \le 2a_1 b_1$;

$= \min\{a_1,b_1\}$ if $\min\{a_1,b_1\} > 2a_1b_1$.

Given (6) and (7), it is clear that the largest value V can assume is given by:

(8) $V^* = 1/a_2b_2$ if $\min\{a_1,b_1\} \leq 2a_1b_1$;
$= [\min\{a_1,b_1\}-a_1b_1]/a_1a_2b_1b_2$ if $\min\{a_1, b_1\} > 2a_1b_1$.

Recall from (3) that the aggregate index of discrimination δ which we are after is just the quantity V/V^*, so that, in view of (5) and (8), we have:

(9) $\delta = |t_{11}-a_1b_1|/a_1b_1$ if $\min\{a_1,b_1\} \leq 2a_1b_1$;
$= |t_{11}-a_1b_1|/[\min\{a_1,b_1\}-a_1b_1]$ if $\min\{a_1,b_1\} > 2a_1b_1$.

(9) can be written more elaborately as follows:

(10) $\delta = |t_{11}-a_1b_1|/a_1b_1$ if $\min\{a_1, b_1\} \leq 2a_1b_1$;
$= |t_{11}-a_1b_1|/a_1b_2$ if $\min\{a_1, b_1\} > 2a_1b_1$ and $\min\{a_1,b_1\} = a_1$;
$= |t_{11}-a_1b_1|/a_2b_1$ if $\min\{a_1, b_1\} > 2a_1b_1$ and $\min\{a_1,b_1\} = b_1$.

In (10), we have a simple expression for the overall extent of discrimination in a society; with a little manipulation, it is very easy to see that δ is just a multiplicative function of δ_{11}. In fact, δ is precisely equal to δ_{11} when $\min\{a_1,b_1\} \leq 2a_1b_1$; δ is equal to $(b_1/b_2)\delta_{11}$ when $a_1 \leq b_1 < 1/2$; and δ is equal to $(a_1/a_2)\delta_{11}$ when $b_1 \leq a_1 < 1/2$. (10) indicates that the only data that need to be accessed in order to compute the value of δ are (i) the proportion of each identity category's population in the total population; (ii) the proportion of each well-being category's population in the total population; and (iii) for each identity category, the proportion of its population in each of the well-being categories. The index δ, being a normalised index, has the convenient property of lying in the closed interval [0,1], with a higher value of δ signifying a relatively greater degree of discrimination.

Well-Being and Caste-Discrimination in India: Some Illustrative Examples

We now consider a few examples of caste-related discrimination from Indian data, the object being simply to provide some empirical applications for the measurement problem reviewed above[2]. Specifically, we shall look for evidence on the extent to which the Scheduled Castes and Tribes are relatively disadvantaged in the matter of access to income, knowledge, government jobs, and some basic amenities of housing infrastructure. In each case we shall present the data required to compute the discrimination index δ in the form of the 2x2 $[t_{ij}]$ matrix in which, to recall, t_{ij} is the proportion of the population that belongs to the ith identity category (i = 1, 2) and the jth well-being category (j = 1, 2). In every case, identity category 1 will be taken to be represented by the Scheduled Castes and Scheduled Tribes (SCST) group, and identity category 2 by the rest of the population, designated simply as the 'Others' group. In the case of *income*, well-being category 1 will be represented by the group of *poor* persons, and well-being category 2 by the group of *nonpoor* persons. In the case of *knowledge*, well-being category 1 will be represented by those who are *illiterate* and well-being category 2 by those who are *literate*. In the case of *Central Government jobs*, well-being category 1 will be constituted by the set of *Class III and Class IV employees,* and well-being category 2 by the set of *Class I employees*. Finally, in the case of *housing amenity*, well-being category 1 will be taken to be represented by the set of households living in houses *without electricity*, and well-being category 2 by the set of households dwelling in houses *with electricity*. A typical [t_{ij}] matrix will look as follows.

A Typical [t_{ij}] Matrix

Identity Category / Well-being category	Well-being category 1	Well-being category 2	Row Total
Identity Category 1	t_{11}	t_{12}	a_1
Identity Category 2	t_{21}	t_{22}	a_2
Column Total	b_1	b_2	1

We consider differential access to income first. National Sample Survey (NSS) data for 1983 are available on the distribution of consumption expenditure across expenditure size-classes, separately for the Scheduled Castes, the Scheduled Tribes, and the population as a whole. Jayaraj and Subramanian (1999) have used these data to present groupwise poverty profiles for rural India: they have clubbed the Scheduled Caste and Scheduled Tribe populations into a composite SCST group, and treated the rest of the population as constituting a composite 'Others' group. Employing a rural poverty line of Rs.15 per person per month at 1960–61 prices, and updating this poverty line to current prices by using the Consumer Price Index of Agricultural Labourers (Food), the authors have computed the 1983 rural headcount ratios of poverty separately for the SCST and the 'Others' groups. These estimates for rural India in 1983 suggest that the SCST group accounted for 28.76 per cent of the total population; that 49.58 per cent of the SCST group were in poverty; and that the poor accounted for 36.04 per cent of the total population. With this information, it is possible to generate the relevant [t_{ij}] matrix, as is done in Table 1.

Table 1: The [t_{ij}] Matrix for Caste and Income Status— Rural India, 1983

Identity Category	Well-being Category Poor	Nonpoor	Row Total
SCST	0.1426	0.1450	0.2876
Others	0.2178	0.4946	0.7124
Column Total	0.3604	0.6396	1.0000

Source: Computed from data in National Sample Survey Organisation (1986): *Pattern of Consumer Expenditure of Scheduled Caste and Scheduled Tribe Households*, Report No.332, Government of India.

Next, in the dimension of *knowledge*, Census data for 1991 suggest that the (age 7+) SCST population accounted for 23.66 per cent of the total (7+) population; that 64.65 per cent of the (7+) SCST population was illiterate; and that the (7+) illiteracy rate for the population as a whole was 47.80 per cent. Table 2, based on this information, presents the corresponding [t_{ij}] matrix.

Table 2: The [t_{ij}] Matrix for Caste and (age 7+) Literacy Status— India 1991

Identity Category	Well-being Category Illiterate	Literate	Row Total
SCST	0.1530	0.0836	0.2366
Others	0.3250	0.4384	0.7634
Column Total	0.4780	0.5220	1.0000

Source: Based on data in (i) Census of India, 1991, Series 1, Final Population Tables: *Brief Analysis of Primary Census Abstract (India)*; and (ii) Census of India 1991, Series 1, Paper-1, *Union Primary Census Abstracts for Scheduled Castes and Scheduled Tribes*.

In the matter of *Central Government jobs*, information available for 1980 in the *Report of the Backward Classes Commission* (under the Chairmanship of B.P. Mandal) indicates the following. If we restrict attention to the population constituted by employees in Service Classes I, III, and IV, it turns out that the SCST group accounts for 21.53 per cent of the population; that 86.77 per cent of the SCST group is employed in Classes III and IV; and that Classes III and IV employees together account for 60.36 per cent of all employees. The [t_{ij}] matrix compatible with these data is presented in Table 3.

Table 3: The $[t_{ij}]$ Matrix for Caste and Central Government Jobs — India 1980

Identity Category	Well-being Category: Classes III & IV Service	Class I Service	Row Total
SCST	0.1865	0.0285	0.2153
Others	0.4168	0.3679	0.7847
Column Total	0.6036	0.3964	1.0000

Source: Based on data in Statement No.1 of Annexure VIII, Vol.1: *Report of the Backward Classes Commission* (Government of India 1980).

Finally, data on the caste-wise distribution of households living in houses with and without electricity are available in the 1991 Census of India. These data for 1991 suggest that SCST households accounted for 25.88 per cent of all households; that of the SCST households, 73.65 per cent lived in households without electricity; and that 57.56 per cent of all households lived in houses without electricity. Using these data, one can generate the appropriate $[t_{ij}]$ matrix, as in Table 4.

Table 4: The $[t_{ij}]$ Matrix for Caste and Housing Amenity Status — India 1991

Identity Category	Well-being Category: HHs in houses w/o electricity	HHs in houses with electricity	Row Total
SCST	0.1906	0.0682	0.2588
Others	0.3850	0.3562	0.7412
Column Total	0.5756	0.4244	1.0000

Source: Based on data in (i) Census of India, Series 1, Paper-1, *Union Primary Census Abstracts for Scheduled Castes and Scheduled Tribes*; and (ii) Census of India, 1991, Paper-2 of 1993, *Housing and Amenities: A Brief Analysis of the Housing Tables of 1991 Census*.

Given the information provided in Tables 1–4, it is a simple matter, employing the expression provided in equation (10) of Section 2, to compute the index of discrimination δ for each of the four cases under review. The relevant estimates are summarised in Table 5.

Table 5: The Index δ of Caste Discrimination in Access to Selected Dimensions of Well-Being in India

Welfare Dimension	Index of Caste Discrimination δ
Caste and Income Status (1983)	0.2114
Caste and Literacy Status (1991)	0.3231
Caste and Government Job Status (1980)	0.4346
Caste and Housing Amenity (Electricity) Status (1991)	0.2792

Source: Calculations based on information provided in the $[t_{ij}]$ matrices of Tables 1–4.

Table 5 does not present an attractive picture. Any further comment would amount to gilding the lily — a sentiment that is well warranted even if, given the numbers in Table 5, it would be hard to find a more inappropriate metaphor than this for the context at hand.

Concluding Observations

Disadvantage is frequently a manifestation of the violation of both positive and negative freedoms for the affected agent. Pernicious as these unfreedoms are, they acquire an additional, and particularly raw, edge when they are mediated by discriminatory barriers to entry erected against identifiable sections of the population. This note has been concerned with the modest exercise of deriving an elementary index of group-related discrimination. A few empirical applications of the note's primary concern — that with measurement — have been provided. These applications are intended to serve as mere illustrative examples. Even so, they point to a picture of generalised deprivation on the well-being front, a picture that becomes markedly worse when account is taken of caste-related disabilities: high levels of relative disadvantage against a background of low average achievement are especially distressing and unacceptable. To the extent that quantification has any value, there is a case for a systematic presentation of summary statistics on both deprivation and group discrimination in India. This note is a small methodological contribution to that possible end.

Notes

1. The general approach to reckoning discrimination that has been adopted here is motivationally similar in spirit to the approach adopted in Ramachandran (1990), whose indices of ‘access’ and ‘mobility’ (attributed to K. Nagaraj) are of relevance in the present context. In this connection, the reader is also referred to Radhakrishnan (1986).
2. For a review of secondary data revealing the differential status experienced by the Scheduled Castes and Tribes in relation to occupational and demographic characteristics, see Thangaraj (1995).

References

Jayaraj, D., and S. Subramanian (1999). Poverty and Discrimination: Measurement and Evidence from Rural India. In Barbara Harriss-White and S. Subramanian (eds.). *Illfare in India: Essays on India’s Social Sector in Honour of S. Guhan.* New Delhi: Sage Publications.

Radhakrishnan, P. (1986). *Caste Discrimination in Indigenous Indian Education. III — Study of Early Nineteenth Century Bombay Presidency.* Madras Institute of Development Studies, Working Paper 64, Chennai.

Ramachandran, V.K. (1990). *Wage Labour and Unfreedom in Agriculture: An Indian Case Study.* Clarendon Press, Oxford.

Thangaraj, M. (1995). *Demographic and Occupational Characteristics of the Scheduled Castes and Scheduled Tribes in India*. Madras Institute of Development Studies Working Paper 134, Chennai.