

①

contains Refinement Theory, General Df of Correlation
and General Invariance Theorem.

5 Continuous Distributions:

redo

Everything that has been said so far has been restricted to discrete probability distributions. We turn now to consider the consequences of assuming continuous distributions.

Suppose that we have a ~~distribution~~ density function $P(x)$ over the single random variable X , which now takes its values in the real numbers. Then $P(x)dx$ represents the probability that X lies within $(x, x+dx)$.

If one imagines the range of X to be broken into discrete intervals and is interested only in which interval X will be found, then we are concerned with a discrete distribution. The probability P_i that X will be found in the i^{th} interval (a_i, a_{i+1}) is $P_i = \int_{a_i}^{a_{i+1}} P(x)dx$, a distribution which leads to an information $I = \sum_i P_i \ln P_i$, which is, however, highly sensitive to the manner of choosing the intervals.

Let us suppose that we divide the range into uniformly spaced intervals, of length l , so that the prob of the N^{th} interval is $P_n = \int_{ml}^{(m+1)l} P(x)dx$ (m integer from - ∞ to ∞)

$$\text{then } I_l = \sum_{m=-\infty}^{\infty} P_n \ln P_n = \sum_{m=-\infty}^{\infty} \int_{ml}^{(m+1)l} P(x)dx \ln \int_{ml}^{(m+1)l} P(x)dx$$

and we notice that as we refine our subdivision, as $l \rightarrow 0$, that I_l diverges to $-\infty$, since for sufficiently small l $P_n = \int_{ml}^{(m+1)l} P(x)dx \sim l P(ml)$

$$\Rightarrow I_l \sim \sum_{n=-\infty}^{\infty} (P_n) \ln l P(nl) = \sum_{n=-\infty}^{\infty} P_n \ln P(nl) + \ln l$$

$$n l \sum_{n=-\infty}^{\infty} P(nl) \ln P(nl) + \ln l$$

$$\sum_{n=-\infty}^{\infty} (n l)^2 P(nl) \ln P(nl)$$

while $\ln l \rightarrow -\infty$. On the other hand, if we define $\tilde{I}_x = I_x - \ln l$, so to speak correcting for the effects of subdivision at each stage, we obviously have

$$\lim_{l \rightarrow \infty} \tilde{I}_x = \int p(x) \ln p(x) dx = \tilde{I}_x$$

which we shall define as the information of the continuous distribution $p(x)$. While this definition may seem somewhat arbitrary, we would like to point out that information itself is rarely important, but only comparisons, or differences of information between two distributions have real significance. Thus, if we have two distributions $p_1(x)$ and $p_2(x)$

$$\begin{aligned} \tilde{I}_x^1 - \tilde{I}_x^2 &= \lim_{l \rightarrow \infty} [\tilde{I}_x^1 - \tilde{I}_x^2] = \lim_{l \rightarrow \infty} [I_x^1 - \ln l - (I_x^2 - \ln l)] \\ &= \lim_{l \rightarrow \infty} [I_x^1 - I_x^2] \end{aligned}$$

So that for comparisons this definition of information gives the same difference of information of two distributions as does the limit of the difference of information of the discrete approximations.

Better approach, Given $P_1^1(x)$, $P_2^2(x)$, subdivide

intervall length l , $P_m^1 = \int_{m\ell}^{(m+1)\ell} P^1(x) dx$ $P_n^2 = \int_{n\ell}^{(n+1)\ell} P^2(x) dx$

$$I_e^1 - I_e^2 = \sum_n (P_m^1 \ln P_m^1) - (P_n^2 \ln P_n^2)$$

$$= \sum_n \left[\int_{m\ell}^{(m+1)\ell} P^1(x) dx \ln \int_{m\ell}^{(m+1)\ell} P^1(x) dx - \int_{n\ell}^{(n+1)\ell} P^2(x) dx \ln \int_{n\ell}^{(n+1)\ell} P^2(x) dx \right]$$

as $l \rightarrow 0$ $\int_{m\ell}^{(m+1)\ell} P(x) dx \rightarrow l P(\xi)$ $m\ell < \xi < l+1$ (mean value thm)

$$\therefore = \sum_n \left[\int_{m\ell}^{(m+1)\ell} P^1(x) dx \ln P^1(\xi_1) - \int_{n\ell}^{(n+1)\ell} P^2(x) dx \ln P^2(\xi_2) \right]$$

~~$$+ \left[\sum_n \int_{n\ell}^{(n+1)\ell} P^1(x) dx \right] \ln l - \left[\sum_n \int_{n\ell}^{(n+1)\ell} P^2(x) dx \right] \ln l$$~~

$$\approx \sum_n \left[\int_{m\ell}^{(m+1)\ell} P^1(x) \ln P^1(x) dx - \int_{n\ell}^{(n+1)\ell} P^2(x) \ln P^2(x) dx \right]$$

$$= \underline{\int P^1(x) \ln P^1(x) dx - \int P^2(x) \ln P^2(x) dx}$$

(can do more rigorously by considering difference of
and discrete approx, showing it goes to zero.)

We wish now to prove a theorem concerning the effect upon the correlation of identifying values of random variables, i.e. of failing to distinguish between two values. Let $P_{ij} = P(X_i, Y_j)$ be a discrete joint distribution, and suppose that we choose to regard the occurrence of X_1 or X_2 as a single event \tilde{X}_0 .

Then our new distribution

$$P'_{ij} = \begin{cases} P_{ij} & \text{for } i \neq 0, 1, 2 \\ P_{1j} + P_{2j} & \text{for } i = 0 \\ \text{undefined (or 0) for } i = 1 \text{ or } 2 \end{cases}$$

$$P'_j = P_j \quad P'_i = P_i \quad i = 3, 4, \dots$$

$$\text{Then} \quad P'_0 = P_1 + P_2$$

$$\{X'\}_{\tilde{Y}} - \{X, Y\} = \sum_{i,j} P'_{ij} \ln \frac{P'_{ij}}{P'_i P'_j} - \sum_{i,j} P_{ij} \ln \frac{P_{ij}}{P_i P_j}$$

$$= \sum_{i=3, j=1} P_{ij} \ln \frac{P_{ij}}{P_i P_j} + \sum_{j=1} P_{0j} \ln \frac{P_{0j}}{(P_1 + P_2) P_j}$$

$$- \sum_{i=3, j=1} P_{ij} \ln \frac{P_{ij}}{P_i P_j} - \sum_j \left[\left(P_{1j} \ln \frac{P_{1j}}{P_1 P_j} + P_{2j} \ln \frac{P_{2j}}{P_2 P_j} \right) \right]$$

$$= \sum_j (P_{1j} + P_{2j}) \ln \frac{(P_{1j} + P_{2j})}{(P_1 + P_2) P_j} - \sum_j P_{1j} \ln \frac{P_{1j}}{P_1 P_j} - \sum_j P_{2j} \ln \frac{P_{2j}}{P_2 P_j}$$

$$\leq 0 \quad \text{by Lemma 1}$$

$$\text{so that } \{X, Y\}' \leq \{X, Y\} \quad \text{and identification reduces correlation.}$$

(maybe better, ... see in it) find X_1 ...

Proof of inequality

$$a_i, b_i > 0$$

Lemma 2 $\left(\sum_i b_i \right) \ln \left(\frac{\sum b_i}{\sum a_i} \right) \leq \sum_i b_i \ln \frac{b_i}{a_i}$

Proof, Let $P_i = \frac{a_i}{\sum a_i}$ so that $P_i \geq 0, \sum P_i = 1$

also $\sum_i P_i \left(\frac{b_i}{a_i} \right) = \sum_i \frac{a_i}{(\sum a_i)} \frac{b_i}{a_i} = \frac{\sum b_i}{\sum a_i}$

But, by Theorem () on prob mixtures (convex)

$$\left[\sum_i P_i \left(\frac{b_i}{a_i} \right) \right] \ln \left[\sum_i P_i \left(\frac{b_i}{a_i} \right) \right] \leq \sum_i P_i \left(\frac{b_i}{a_i} \right) \ln \left(\frac{b_i}{a_i} \right)$$

$$\Rightarrow \left(\frac{\sum b_i}{\sum a_i} \right) \ln \left(\frac{\sum b_i}{\sum a_i} \right) \leq \sum_i \frac{a_i}{\sum a_i} \left(\frac{b_i}{a_i} \right) \ln \left(\frac{b_i}{a_i} \right)$$

$$\Rightarrow \left(\sum_i b_i \right) \ln \left(\frac{\sum b_i}{\sum a_i} \right) \leq \sum_i b_i \ln \left(\frac{b_i}{a_i} \right)$$

QED

(strictly less

unless $\frac{b_i}{a_i} = \frac{b_j}{a_j}$ all j)

And we have proved that any refinement of a discrete joint distribution, by which we mean breaking a value of a random variable into a multiple of values, never decreases the correlation. This circumstance allows us to give a rigorous definition of correlation, which will apply to joint probability distributions over completely arbitrary sets, regardless of cardinality, ie for any probability measure over an arbitrary product space, in the following simple manner.

Given two arbitrary sets X and Y , and a probability measure over the product space $M_p\{X \times Y\}$, we consider a ^(discrete)_{countable} partition P of X into measurable sets X_i , and Y into Y_j , for which M_p leads to a discrete probability distribution, with correlation $\{\{X, Y\}\}^P$. We next consider a sequence of partitions P_1, P_2, \dots such that each succeeding one is a refinement of its predecessor, and such that the marginal probability measure of any subset in a partition goes to zero as $P_n \rightarrow n \rightarrow \infty$.

By Theorem () $\{\{X, Y\}\}^{P_m} \geq \{\{X, Y\}\}^P$ if $m > n$ so

that the sequence $\{\{X, Y\}\}^{P_m}$ is monotone increasing with m , hence always possesses a limit (possibly ∞), which we define as the correlation $\{\{X, Y\}\}$.

$$\text{f) } \{\{X, Y\}\} = \lim_{n \rightarrow \infty} \{\{X, Y\}\}^{P_n} = \sup_P \{\{X, Y\}\}^P$$

(Note) The limit for any sequence "obviously" converges to \uparrow .

(What is true is that for any given P there is a refinement of P , P' such that P' is arbitrarily close to $\sup \{\{X, Y\}\}$.

General Invariance of Correlation:

Let X, Y be sets with probability measure $M_p(X \times Y)$ on the product space.

Let $f(X)$ be a one to one mapping of X onto U
 $g(Y)$ be a one to one mapping of Y onto V

Then the probability measure M' produced on the product space $U \times V$ is $P'(U_i, V_j) = P(f^{-1}(U_i), g^{-1}(V_j))$

Consider any partition P of X, Y into $\{X_i\}, \{Y_j\}$ which has distribution P_{ij} . The corresponding

Partition P' in UV -space $\{U_i\} \{V_j\}$ where $U_i = f(X_i)$ (image of X_i)
 $V_j = g(Y_j)$ (image of Y_j)

Leads to a joint discrete distribution $P'_{ij} = P'(U_i, V_j) = P(X_i, Y_j) = P_{ij}$

that is, a numerically identical distribution, so that

$$\{X, Y\}^{P'} = \{U, V\}^{P'}$$

$$\Rightarrow \sup_P \{X, Y\}^P = \sup_{P'} \{U, V\}^{P'}$$

(or contradiction
since each P has a P'
with $\{X, Y\} = \{U, V\}$)

And we have proved in General:

$$\{X, Y\} = \{U, V\}$$

for U any one-to-one mapping of X
 V any one-to-one mapping of Y
 (over)

This theorem & Proceeding
apply Equally to all the
Correlation Brackets!

(8)

Specializing to continuous real variables x, y ,
with continuous distribution $\rho(x, y)$