



Secretaria de Planejamento e Coordenação da Presidência da República
Fundação Instituto Brasileiro de Geografia e Estatística — IBGE

RELATÓRIO TÉCNICO

ENCE



ESCOLA
NACIONAL
DE CIÊNCIAS
ESTATÍSTICAS

CURVA DE LORENZ E ÍNDICE DE GINI
PARA DISTRIBUIÇÕES DE RENDA

José Paulo Q. Carneiro - IBGE/ENCE
Jorge Luiz Rangel Costa - CIDE

CURVA DE LORENZ E ÍNDICE DE GINI
PARA DISTRIBUIÇÕES DE RENDA

José Paulo Q. Carneiro - IBGE/ENCE
Jorge Luiz Rangel Costa - CIDE

As matérias publicadas nos RELATÓRIOS TÉCNICOS são preprints, com tiragens limitadas, de trabalhos elaborados por professores da ENCE, em complementação a suas atividades de ensino, com ênfase para as pesquisas realizadas no Laboratório de Estatística da Escola.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA — IBGE

ESCOLA NACIONAL DE CIÊNCIAS ESTATÍSTICAS - ENCE

R E L A T Ó R I O S T É C N I C O S

Nº 02/89

CURVA DE LORENZ E ÍNDICE DE GINI
PARA DISTRIBUIÇÕES DE RENDA

José Paulo Q. Carneiro - IBGE/ENCE
Jorge Luiz Rangel Costa - CIDE

RIO DE JANEIRO

Novembro/89

CURVA DE LORENZ E ÍNDICE DE GINI PARA DISTRIBUIÇÕES DE RENDA

José Paulo Q. Carneiro - IBGE / ENCE

Jorge Luiz Rangel Costa - CIDE

1. Introdução

Uma dada distribuição de renda pode ser encarada como qualquer outra distribuição de frequência. Como tal, podem-se aplicar a ela as medidas tradicionais de locação e de dispersão, tais como média, mediana, desvio padrão, diferença inter-quartilica, etc. No entanto, para distribuições de renda, tornou-se tradicional medir sua concentração através do chamado **Índice de Gini**, principalmente por causa de sua interpretação geométrica, relacionada à também tradicional **curva de Lorenz**. O objetivo destas notas é esclarecer, a nível de exposição, alguns aspectos matemáticos do assunto e aplicá-los a dados do Brasil e do Rio de Janeiro.

Ao estudar distribuições de frequência ou de probabilidade, uma atitude possível é tratar separadamente os casos "discreto" e "contínuo". Mais elegante ainda seria um tratamento unificado, através da integral de Stieltjes. Em face, porém, do nível do presente texto e da simplicidade dos modelos e dos casos apreciados, optou-se por fazer a abordagem teórica somente no caso contínuo, utilizando apenas integral de Riemann. Na realidade, vamos supor que todas as funções utilizadas são continuamente diferenciáveis exceto no máximo em um número finito de pontos, e, nestes pontos, possuem limites laterais finitos. O leitor versado em integral de Stieltjes poderá constatar que um tratamento análogo valeria para o caso geral. Mais adiante, alguns parágrafos são dedicados ao importante fato de que as distribuições de renda que se encontram na prática provêm de um número finito de observações e aparecem grupadas por classes.

2. A Diferença média

Para uma função de densidade de probabilidade f , talvez as medidas de dispersão mais populares sejam o desvio padrão σ , cujo quadrado (a variância) é definido como:

$$\sigma^2 = \int_{-\infty}^{+\infty} (x - \mu)^2 f(x) dx \quad (1)$$

e o desvio médio

$$\delta_1 = \int_{-\infty}^{+\infty} |x - \mu| f(x) dx \quad (2)$$

Nessas duas fórmulas, $\mu = \int_{-\infty}^{+\infty} x f(x) dx$ é a média da distribuição.

Procurando medidas que não dependam dos afastamentos em relação a um valor central, mas somente das diferenças entre os próprios valores observados, somos levados a introduzir os análogos:

$$E^2 = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x - y)^2 f(x) f(y) dy dx \quad (3)$$

$$\Delta = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} |x - y| f(x) f(y) dy dx \quad (4)$$

A fórmula (3) não traz muito de novo em relação a (1), já que se verifica facilmente que $E^2 = 2\sigma^2$, utilizando as definições e o fato de que $\int_{-\infty}^{+\infty} f(x) dx = 1$.

Já a medida de dispersão Δ , chamada **diferença média**, tem uma importância especial e entrará na própria definição do índice de Gini. Para expressá-la sem utilizar valor absoluto, decompõe-se a integral "interior" em:

$$\int_{-\infty}^x (x - y) f(x) f(y) dy + \int_x^{+\infty} (y - x) f(x) f(y) dy$$

onde a segunda parcela, por sua vez, pode ser escrita como

$$\int_{-\infty}^{+\infty} - \int_{-\infty}^x$$

Com isto, chega-se à expressão equivalente para Δ :

$$\Delta = 2 \int_{-\infty}^{+\infty} \int_{-\infty}^x (x-y)f(y)f(x)dydx \quad (5)$$

3. Distribuições de Renda

Uma distribuição de renda tem, evidentemente, a peculiaridade de que sua função de densidade de probabilidade $f(x)$ é nula para $x < 0$, o que será suposto de agora por diante. Na prática, ela também será nula para valores acima de uma certa renda máxima; como, porém, quer-se deixar lugar para modelos um pouco mais gerais (Pareto, lognormal, etc.), esta condição não será considerada uma hipótese. Com esses pressupostos e mais os da Introdução, a função de distribuição F associada a f , definida por $F(x) = \int_0^x f$ é tal que $F(0) = 0$ e $F'(x) = f(x)$, exceto no máximo em um número finito de pontos. Além disto, a média μ será positiva, exceto em casos triviais, e aqui será sempre suposta > 0 .

Tudo isto permite deduzir duas outras expressões úteis para Δ . Em primeiro lugar, para x fixo, pondo $u = x - y$, $dv = f(y)dy$, e integrando por partes:

$$\int_0^x (x-y)f(y)dy = \int_0^x F(y)dy$$

Levando este resultado em (5), e trocando, em seguida, a ordem de integração, vem:

$$\begin{aligned} \frac{\Delta}{2} &= \int_0^{+\infty} \int_0^x f(x)F(y)dydx \\ &= \int_0^{+\infty} \int_y^{+\infty} f(x)F(y)dx dy \\ &= \int_0^{+\infty} F(y) \int_y^{+\infty} f(x)dx dy \\ &= \int_0^{+\infty} F(y)(1-F(y)) dy \end{aligned}$$

Portanto:

$$\Delta = 2 \int_0^{+\infty} F(x)(1-F(x)) dx \quad (6)$$

uma expressão que não utiliza integral dupla.

Analogamente, fazendo, em (6), $u = F(x)(1 - F(x))$, $dv = dx$, e integrando por partes, chega-se a:

$$\Delta = 4 \int_0^{+\infty} x \left(F(x) - \frac{1}{2} \right) f(x) dx \quad (7)$$

4. O Índice de Gini

Na expressão (7), como $F(x) \leq 1$, então $F(x) - \frac{1}{2} \leq \frac{1}{2}$. Portanto:

$$\frac{\Delta}{4} \leq \frac{1}{2} \int_0^{+\infty} x f(x) dx = \frac{\mu}{2}$$

Segue-se que

$$0 \leq \frac{\Delta}{2\mu} \leq 1$$

O número

$$G = \frac{\Delta}{2\mu} \quad (8)$$

é, por definição, o **índice de Gini** da distribuição. Ao dividir Δ pelo dobro da média, o índice de Gini relativiza a diferença média, tal como faz o coeficiente de variação com o desvio padrão. Porém o coeficiente de variação não tem a propriedade agradável de estar entre 0 e 1.

5. A Curva de Lorenz

Desde Lorenz (1905), tornou-se habitual, para distribuições de renda, construir uma curva através de pontos (p_i, q_i) , onde p_i representa a proporção de indivíduos que possuem renda inferior a um certo valor r_i , enquanto q_i é a proporção da renda total possuída por esses indivíduos, sendo $r_1 < \dots < r_n$. É a chamada **curva de Lorenz**.

No caso contínuo, esse procedimento equivale ao seguinte: imagine-se que f seja a função de densidade de probabilidade associada a uma distribuição de renda, isto é, $\int_{x_1}^{x_2} f$ mede a probabilidade de que a renda de um indivíduo esteja entre x_1 e x_2 . Neste caso, a proporção da renda total possuída pelas pessoas com renda $\leq x$ é dada por:

$$\Phi(x) = \frac{\int_0^x t f(t) dt}{\int_0^{+\infty} t f(t) dt}$$

Levando em conta que o denominador é a média μ da distribuição, isto é, a renda média geral da população, pode-se escrever:

$$\Phi(x) = \frac{1}{\mu} \int_0^x t f(t) dt$$

Como $F'(x) = f(x) \geq 0$ e $\Phi'(x) = \frac{x f(x)}{\mu} \geq 0$, as funções F e Φ são ambas crescentes, e portanto inversíveis, no intervalo $[a; w]$, onde a e w são as rendas mínima e máxima, ou $w = +\infty$, no caso de modelos sem renda máxima.

Considerando x em $[a; w]$ como um parâmetro, então a curva $v = L(u)$ dada pelas equações paramétricas

$$\begin{cases} u = F(x) \\ v = \Phi(x) \end{cases}$$

é a curva de Lorenz associada à distribuição definida por f . Como $v = \Phi(F^{-1}(u))$, vê-se que $L = \Phi \circ F^{-1}$, e como u e v variam entre 0 e 1, a curva se encontra dentro do quadrado unitário.

Examinemos as derivadas.

$$\frac{dv}{du} = \frac{dv/dx}{du/dx} = \frac{\Phi'(x)}{F'(x)} = \frac{x f(x)/\mu}{f(x)} = \frac{x}{\mu} > 0$$

já que $x > 0$. Logo a curva de Lorenz é estritamente crescente.

$$\frac{d^2v}{du^2} = \frac{d(\frac{dv}{du})/dx}{du/dx} = \frac{1/\mu}{f(x)} > 0$$

para $a < x < w$. Logo a curva de Lorenz é convexa e portanto o gráfico está sempre abaixo da secante $v = u$ e acima de suas tangentes.

6. A Discrepância máxima

Para cada $0 \leq u \leq 1$ a diferença $u - v = u - L(u) = g(u)$ mede a distância vertical entre a reta $v = u$ e a curva de Lorenz. Como $g'(u) = 1 - \frac{x}{\mu}$, vê-se que $g'(u) > 0$ para $x < \mu$ e $g'(u) < 0$ para $x > \mu$. Portanto $g(u)$ é máxima para $x = \mu$, e, neste ponto, $\frac{dv}{du} = 1$, isto é, a tangente à curva de Lorenz é paralela à reta

$v = u$. Isto significa que quando se diz: "os $u\%$ mais pobres possuem $v\%$ da renda total", a maior diferença $u - v$ ocorrerá quando se estiver referindo àqueles que ganham até a renda média da distribuição.

O número $\delta = F(\mu) - \Phi(\mu)$, valor máximo de $g(u)$, é chamado **discrepância máxima**.

A discrepância máxima é também uma medida de dispersão relativa, já que:

$$\delta = \frac{\delta_1}{2\mu} \quad (9)$$

onde δ_1 é o desvio médio da distribuição, definido em (2). Isto se demonstra imediatamente, decompondo a integral da definição (2) em :

$$\delta_1 = \int_0^{\mu} (\mu - x)f(x)dx + \int_{\mu}^{+\infty} (x - \mu)f(x)dx.$$

Em seguida, observa-se que :

$$\int_0^{\mu} (x - \mu)f(x)dx + \int_{\mu}^{+\infty} (x - \mu)f(x)dx = \int_0^{+\infty} (x - \mu)f(x)dx = 0$$

concluindo-se pois que $\delta_1 = 2 \int_0^{\mu} (\mu - x)f(x)dx$.

Por outro lado, pelas definições de F e Φ ,

$$\delta = F(\mu) - \Phi(\mu) = \frac{1}{\mu} \int_0^{\mu} (\mu - x)f(x)dx = \frac{\delta_1}{2\mu}.$$

7. Interpretações Geométricas

A Figura 1 apresenta, no plano uv , uma curva de Lorenz $v = L(u)$, com as propriedades geométricas deduzidas no parágrafo 5.

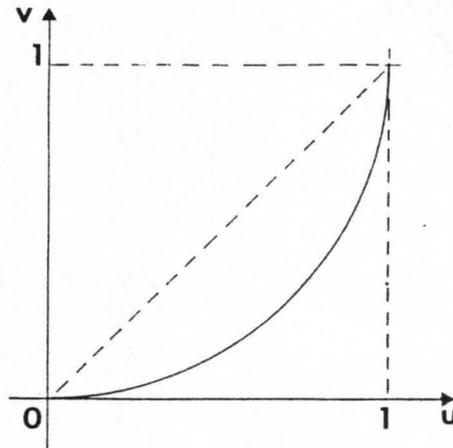


Figura 1 - Curva de Lorenz

Sejam então:

$S =$ área da região definida por: $0 \leq u \leq 1; L(u) \leq v \leq u$.

$A = \int_0^1 u dv =$ área da região definida por: $0 \leq v \leq 1; 0 \leq u \leq L^{-1}(v)$.

$B = \int_0^1 v du =$ área da região definida por: $0 \leq u \leq 1; 0 \leq v \leq L(u)$.

$C =$ metade da área do quadrado $0 \leq u \leq 1; 0 \leq v \leq 1$.

Tem-se claramente $S = A - C = C - B$, de onde se conclui que

$$2S = A - B = \int_0^1 u dv - \int_0^1 v du$$

Como $u = F(x)$ $v = \Phi(x)$ $du = f(x)dx$ $dv = \frac{x f(x)}{\mu} dx$,
então:

$$\begin{aligned} 2S &= \frac{1}{\mu} \int_0^{+\infty} (xF(x) - \mu\Phi(x)) f(x) dx \\ &= \frac{1}{\mu} \int_0^{+\infty} \left(x \int_0^x f(y) dy - \int_0^x y f(y) \right) f(x) dx \\ &= \frac{1}{\mu} \int_0^{+\infty} \int_0^x (x - y) f(y) dy f(x) dx = \frac{\Delta}{2\mu} = G \end{aligned}$$

pelas fórmulas (5) e (8).

Portanto, o índice de Gini tem como interpretação geométrica o dobro da área da região compreendida entre a curva de Lorenz e a bissetriz $v = u$. Esta interpretação, muito usada no estudo de distribuições de renda, permite verificar graficamente o fato, já demonstrado analiticamente, de que o índice de Gini está entre 0 e 1, já que a área S não pode ultrapassar a metade da área do quadrado, isto é, $1/2$.

Na Figura 2, além da curva de Lorenz, está também representada a discrepância máxima δ , pelo segmento PQ . Pela convexidade da curva de Lorenz, é claro que a área S acima referida está entre a área do triângulo OPR e a área do paralelogramo $ORVT$. Esta última é igual a $PQ \times 1 = \delta$ e também igual ao dobro da área do triângulo. Segue-se que $\frac{\delta}{2} \leq S \leq \delta$ e portanto que:

$$\delta \leq G \leq 2\delta \quad (10)$$

uma desigualdade útil, quando for fácil calcular a discrepância máxima. Além disto, esta fórmula, utilizando (8) e (9), dá também uma relação interessante entre a diferença média e o desvio médio, a saber:

$$\delta_1 \leq \Delta \leq 2\delta_1 \quad (11)$$

É claro que estas fórmulas podem ser deduzidas analiticamente, mas é notável a facilidade com que se as deduz geometricamente.

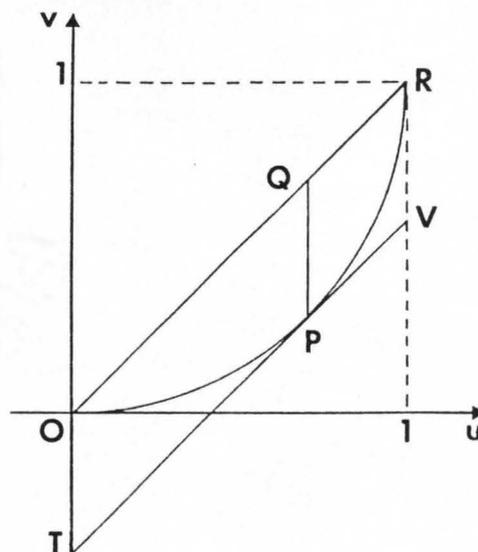


Figura 2 - Discrepância Máxima

8. Curva de Lorenz, Índice de Gini e Outras Características para algumas distribuições

O Quadro 1 apresenta a curva de Lorenz, o índice de Gini e outras características das principais distribuições teóricas de renda. Na realidade, a distribuição "potência" não é encontrada usualmente neste contexto, mas ela foi colocada aqui por permitir algumas especulações interessantes.

Observe-se que:

(a) Podem-se obter índices de Gini tão próximos de zero quanto se desejar, na distribuição uniforme, fazendo $b \rightarrow a$, ou na distribuição lognormal, fazendo $\sigma \rightarrow 0$. Nesses casos, a curva de Lorenz aproxima-se da reta $v = u$. Na realidade, pode-se obter o índice de Gini exatamente igual a zero com a distribuição "concentrada" definida por $F(x) = 1$ para $x = a$ e $F(x) = 0$ para $x \neq a$, mas, para isto, seria necessário utilizar integral de Stieltjes. Naturalmente, este modelo corresponderia ao caso em que todos os indivíduos da população teriam a mesma renda, ou seja, traduziria a mais perfeita homogeneidade.

(b) Podem-se obter índices de Gini tão próximos de 1 quanto se desejar, na distribuição de Pareto, fazendo $p \rightarrow 1^+$, ou na distribuição lognormal, fazendo $\sigma \rightarrow +\infty$. Nestes casos, a curva de Lorenz aproxima-se da reta $v = 0$.

(c) As seguintes três distribuições: uniforme com $b = 2a$, Pareto com $p = 5$ e potência com $p = 4$, têm o mesmo índice de Gini (e as duas últimas, a mesma discrepância máxima). No entanto são completamente diversas em termos de distribuição de renda, como se pode ver em um gráfico da função de densidade de cada uma delas. Isto mostra como o índice de Gini ou a discrepância máxima devem ser utilizados com cuidado.

A dedução das fórmulas do Quadro 1 é, de um modo geral, imediata, a partir das fórmulas apresentadas anteriormente (no caso do índice de Gini, aconselha-se usar a fórmula (6)). A exceção fica por conta da lognormal, que é tratada no Apêndice.

Quadro 1
Curva de Lorenz, Índice de Gini e outras características
para algumas distribuições

	Uniforme	Exponencial	Pareto	Potência	Lognormal
$f(x)$	$\frac{1}{b-a}$	ce^{-cx}	$\frac{pa^p}{ax^{p+1}}$	$\frac{px^{p-1}}{b^p}$	$\frac{1}{\sigma x \sqrt{2\pi}} e^{-\frac{1}{2}(\frac{\log x - \mu}{\sigma})^2}$
Intervalo fora do qual $f(x)=0$	$[a; b]$	$[0; +\infty[$	$[a; +\infty[$	$[0; b]$	$]0; +\infty[$
Condições	$0 \leq a < b$	$c > 0$	$p > 1$ $a > 0$	$p > 1$ $b > 0$	$\sigma > 0$
Curva de Lorenz $L(u)$	$\frac{(b-a)u^2 + 2au}{b+a}$	$u + w \log w$	$1 - w^{\frac{p-1}{p}}$	$u^{\frac{p+1}{p}}$	$N(N^{-1}(u) - \sigma)$
Média	$\frac{a+b}{2}$	$\frac{1}{c}$	$\frac{p}{p-1}a$	$\frac{p}{p+1}b$	$e^{\mu + \frac{\sigma^2}{2}}$
Coefficiente de variação	$\frac{1}{\sqrt{3}} \frac{b-a}{b+a}$	1	$\frac{1}{\sqrt{p(p-2)}}$	$\frac{1}{\sqrt{p(p+2)}}$	$\sqrt{e^{\sigma^2} - 1}$
Índice de Gini	$\frac{1}{3} \frac{b-a}{b+a}$	$\frac{1}{2}$	$\frac{1}{2p-1}$	$\frac{1}{2p+1}$	$2N(\frac{\sigma}{\sqrt{2}}) - 1$
Discrepância máxima	$\frac{1}{4} \frac{b-a}{b+a}$	$\frac{1}{6}$	$\frac{(p-1)^{p-1}}{p^p}$	$\frac{p^p}{(p+1)^{p+1}}$	$2N(\frac{\sigma}{2}) - 1$

Obs.1: Nas equações das curvas de Lorenz, $w = 1 - u$.

Obs.2: O coeficiente de variação da Pareto só existe para $p > 2$.

Obs.3: Na lognormal, $N(z) = N_{0,1}(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-\frac{t^2}{2}} dt$.

9. Cotas para o Índice de Gini com Dados Grupados

Quando se considera, na prática, uma distribuição de renda, os dados aparecem grupados em classes. O grupamento pode ser um "dado", por exemplo, numa tabela de uma publicação, ou pode ser arbitrado pelo pesquisador, se ele estiver confeccionando uma tabulação a partir de dados individuais. Aqui vai ser considerado o primeiro caso. De qualquer modo, o fato de se gruparem os dados em classes evidentemente mascara o índice de Gini,

que é uma medida de dispersão. O que vai ser mostrado agora é que: se a renda média (ou, equivalentemente, a renda total) de cada classe for conhecida, e se se supuser que os dados provêm de uma distribuição não especificada, porém de forma que a curva de Lorenz tenha as propriedades apresentadas no parágrafo 5, particularmente a convexidade, então é possível determinar uma cota inferior e uma cota superior para o índice de Gini, a partir dos dados empíricos. Os exemplos do parágrafo 12 darão uma idéia de quão estreitas são essas cotas. As fórmulas podem ser deduzidas de modo puramente analítico, como em [1], mas aqui serão obtidas geometricamente.

Sejam então:

$0 = r_0 < r_1 < \dots < r_n$ os limites das classes de renda;

$p_i =$ número de pessoas com renda em $[r_{i-1}; r_i[$, para $i = 1, \dots, n$

$p_{n+1} =$ número de pessoas com renda $> r_n$

$P = \sum_{i=1}^{n+1} p_i =$ total de pessoas

$u_i = \frac{\sum_{j=1}^i p_j}{P} =$ proporção de pessoas com renda $< r_i$, para $i = 1, \dots, n+1$

$u_0 = 0$

$\mu_n =$ renda média das pessoas com renda em $[r_{i-1}; r_i[$, para $i = 1, \dots, n$

$\mu_{n+1} =$ renda média das pessoas com renda $> r_n$

(Para facilitar o tratamento da última classe, será criada uma renda fictícia máxima r_{n+1} , determinada por alguma hipótese compatível com a renda média da classe aberta, por exemplo, $r_{n+1} = 2\mu_{n+1} - r_n$).

$p_i \mu_i =$ renda da classe $[r_{i-1}; r_i[$, para $i = 1, \dots, n+1$

$R = \sum_{i=1}^{n+1} p_i \mu_i =$ renda total

$v_i = \frac{1}{R} \sum_{j=1}^i p_j \mu_j =$ proporção da renda total possuída pelas pessoas com renda $< r_i$, para $i = 1, \dots, n+1$.

A Figura 3 mostra os pontos provenientes da distribuição empírica e a curva de Lorenz associada. Pelo parágrafo 7 e pela convexidade da curva de Lorenz, uma cota inferior G_I

para o índice de Gini é dada pelo dobro da área da figura formada pela reta $v = u$ e pela poligonal que une os pontos $(u_i; v_i)$. Logo:

$$G_I = 1 - \sum_{i=1}^{n+1} (u_i - u_{i-1})(v_i + v_{i-1}) \quad (12)$$

Levando em conta o significado de u_i e v_i , a fórmula também pode ser escrita:

$$G_I = 1 - \frac{1}{PR} \left(\sum_{i=1}^{n+1} p_i^2 \mu_i + 2 \sum_{j=1}^{i-1} p_i p_j \mu_j \right) \quad (13)$$

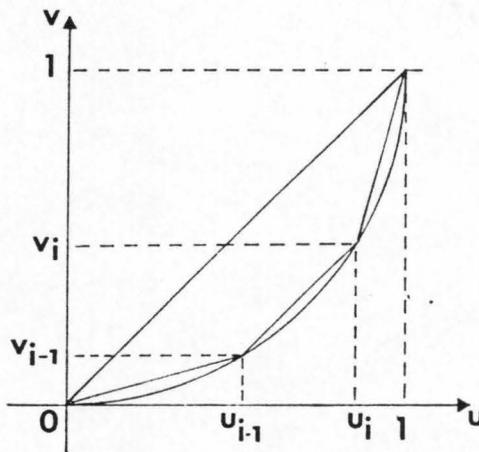


Figura 3 - Cota inferior para o Gini

Ainda pela convexidade, uma cota superior para o índice de Gini será dada pelo dobro da área da figura formada pela reta $v = u$ e pelas tangentes à curva de Lorenz nos pontos $(u_i; v_i)$. Na realidade, essa cota superior G_S pode ser escrita como $G_S = G_I + D$, onde D é o dobro da soma das áreas dos triângulos indicados na Figura 4, formados pelos pontos $(u_{i-1}; v_{i-1})$, $(u_i; v_i)$ e $(\bar{u}_i; \bar{v}_i)$, onde $(\bar{u}_i; \bar{v}_i)$ é o ponto de interseção das tangentes à curva de Lorenz nos pontos $(u_{i-1}; v_{i-1})$ e $(u_i; v_i)$.

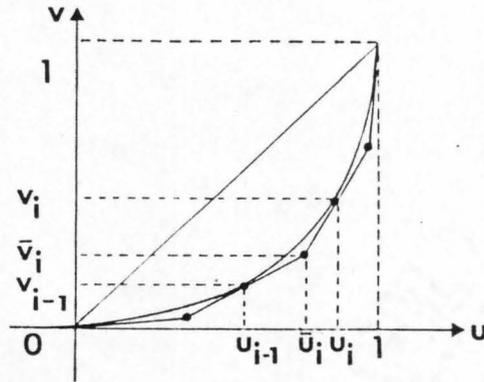


Figura 4 - Cota superior para o Gini

A inclinação da tangente à curva de Lorenz em $(u_i; v_i)$ é dada por $\frac{dv}{du}$, para $u = u_i$. Porém, pelo parágrafo 5, tem-se que $\frac{dv}{du} = \frac{x}{\mu} = \frac{F^{-1}(u)}{\mu}$, onde $\mu = \sum_{i=1}^{n+1} p_i \mu_i$ é a média da distribuição, isto é, a renda média total, e F acumula as pessoas. Logo, a inclinação procurada é $\frac{F^{-1}(u_i)}{\mu} = \frac{r_i}{\mu}$.

As duas tangentes consecutivas têm, portanto, equações:

$$v - v_{i-1} = \frac{r_{i-1}}{\mu}(u - u_{i-1})$$

$$v - v_i = \frac{r_i}{\mu}(u - u_i)$$

Resolvendo o sistema e levando em conta as expressões de u_i e v_i , obtêm-se para o ponto de interseção:

$$\bar{u}_i = (1 - \alpha_i)u_{i-1} + \alpha_i u_i$$

$$\bar{v}_i = (1 - \beta_i)v_{i-1} + \beta_i v_i$$

onde $\alpha_i = \frac{r_i - \mu_i}{r_i - r_{i-1}}$ e $\beta_i = \frac{r_{i-1}}{\mu_i} \alpha_i$.

Finalmente, o dobro da área S_i do triângulo em questão pode ser calculado pela conhecida fórmula do determinante, isto é:

$$2S_i = (\bar{u}_i - u_{i-1})(v_i - v_{i-1}) - (u_i - u_{i-1})(\bar{v}_i - v_{i-1}) = (\alpha_i - \beta_i)(u_i - u_{i-1})(v_i - v_{i-1})$$

Ou seja:

$$2S_i = \left(\frac{p_i}{P}\right)^2 \frac{(\mu_i - r_{i-1})(r_i - \mu_i)}{\mu(r_i - r_{i-1})}$$

Portanto:

$$D = \frac{1}{\mu} \sum_{i=1}^{n+1} \left(\frac{p_i}{P}\right)^2 \frac{(\mu_i - r_{i-1})(r_i - \mu_i)}{r_i - r_{i-1}} \quad (14)$$

A conclusão é que o índice de Gini G se situa entre os valores:

$$G_I \leq G \leq G_I + D = G_S \quad (15)$$

dados pelas fórmulas (13) e (14) .

10. Interpretação da Cota Inferior para o Índice de Gini

Frequentemente a fórmula (12) é apresentada como sendo “a fórmula” do índice de Gini para dados grupados. No entanto, como ficou claro no parágrafo 9, ela fornece apenas uma cota inferior para o índice.

Cabe aqui a questão: fixadas, como no início do parágrafo 9, as classes de renda definidas pelos r_i , os números de pessoas p_i e as rendas médias μ_i de cada classe, existe uma situação correspondente a G_I ?

Como o índice de Gini é uma medida de dispersão, é natural que ele seja mínimo na situação de maior uniformidade, isto é, quando as p_i pessoas com renda entre r_{i-1} e r_i tiverem todas renda igual a μ_i . Vejamos que isto de fato ocorre, em um certo sentido. Para tanto, vamos ter que nos afastar do caso contínuo.

Consideremos então a distribuição discreta correspondente ao caso em que, para cada $i = 1, \dots, n+1$, haja p_i pessoas com renda μ_i . As fórmulas (4) e (5) são, na realidade, casos particulares (no caso contínuo) das fórmulas mais gerais:

$$\Delta = \int_{-\infty}^{+\infty} |x - y| dF(y) dF(x) \quad (16)$$

$$\Delta = 2 \int_{-\infty}^x |x - y| dF(y) dF(x) \quad (17)$$

onde as integrais devem ser entendidas como integrais de Riemann-Stieltjes.

Traduzindo a fórmula (17) para o caso discreto em questão, levando em conta a definição (8), observando que $\mu_{i-1} < \mu_i$, e que $\mu = \frac{R}{P}$, obtêm-se, para o índice de Gini G_D :

$$G_D = \frac{1}{PR} \sum_{i=1}^{n+1} \sum_{j<i}^{n+1} p_i p_j (\mu_i - \mu_j) \quad (18)$$

Por outro lado:

$$\begin{aligned} PR &= \sum_{i=1}^{n+1} p_i \sum_{j=1}^{n+1} p_j \mu_j \\ &= \sum_{i=1}^{n+1} p_i^2 \mu_i + \sum_{i=1}^{n+1} \sum_{j<i}^{n+1} p_i p_j (\mu_i + \mu_j) \\ &= \sum_{i=1}^{n+1} p_i^2 \mu_i + \sum_{i=1}^{n+1} \sum_{j<i}^{n+1} 2p_i p_j \mu_j + \sum_{i=1}^{n+1} \sum_{j<i}^{n+1} p_i p_j (\mu_i - \mu_j) \\ &= PR(1 - G_I) + PRG_D \end{aligned}$$

a última passagem vindo de (13) e (18).

Dai se conclui que $G_D = G_I$.

Resumindo: dada uma distribuição do tipo contínuo, da qual se conhecem os dados apenas em um grupamento em classes fixado, isto é, não se conhece a sua distribuição dentro das classes, então a fórmula (13) fornece o menor valor possível para o índice de Gini, e este valor equivale à hipótese de que dentro das classes todas as pessoas têm a mesma renda.

11. Interpretação da Cota Superior para o Índice de Gini

Imaginemos agora a situação em que as p_i pessoas com renda no intervalo $[r_{i-1}; r_i[$ estejam distribuídas de forma que $\beta_i p_i$ pessoas tenham renda r_{i-1} , enquanto $(1 - \beta_i) p_i$ tenham renda $r_i - \epsilon$, onde $0 < \epsilon < r_i - r_{i-1}$ e $0 < \beta_i < 1$. Isto é, um conjunto de pessoas concentra a renda r_{i-1} e as restantes concentram a renda $r_i - \epsilon$, "próxima" de r_i . Vamos mostrar que o índice de Gini G_ϵ , correspondente a essa situação, tende à cota superior G_S , quando $\epsilon \rightarrow 0$.

Em primeiro lugar, para que a renda média dessas pessoas permaneça igual a μ_i , é necessário que:

$$\beta_i p_i r_{i-1} + (1 - \beta_i) p_i (r_i - \epsilon) = \mu_i$$

Daf se deduz que $\beta_i = \frac{r_i - \mu_i - \epsilon}{r_i - r_{i-1}}$. Note-se que :

$$\lim_{\epsilon \rightarrow 0} \beta_i = \alpha_i = \frac{r_i - \mu_i}{r_i - r_{i-1}}$$

e que:

$$\alpha_i r_{i-1} + (1 - \alpha_i) r_i = \mu_i \quad (19)$$

Para calcular, por uma fórmula análoga à (18), o índice de Gini G_ϵ correspondente ao caso descrito, é preciso observar que há agora $2n + 2$ valores de renda s_k , com $k = 0, \dots, 2n + 1$, de tal forma que, para $j = 0, \dots, n$:

$$s_{2j} = r_j$$

$$s_{2j+1} = r_{j+1} - \epsilon$$

As frequências absolutas associadas a esses valores de renda são q_k , com $k = 0, \dots, 2n + 1$, de modo que, para $j = 0, \dots, n$:

$$q_{2j} = \beta_{j+1} p_{j+1}$$

$$q_{2j+1} = (1 - \beta_{j+1}) p_{j+1}$$

É de notar também que, quando $\epsilon \rightarrow 0$:

$$s_{2j+1} \rightarrow r_{j+1} \quad q_{2j} \rightarrow \alpha_{j+1} p_{j+1} \quad q_{2j+1} \rightarrow (1 - \alpha_{j+1}) p_{j+1} \quad (20)$$

Além disto:

$$\sum_{k=0}^{2n+1} q_k = \sum_{j=0}^n (q_{2j} + q_{2j+1}) = \sum_{j=0}^n p_{j+1} = P$$

$$\begin{aligned} \sum_{k=0}^{2n+1} q_k s_k &= \sum_{j=0}^n (q_{2j} s_{2j} + q_{2j+1} s_{2j+1}) = \sum_{j=0}^n p_{j+1} (\beta_{j+1} r_j + (1 - \beta_{j+1})(r_{j+1} - \epsilon)) = \\ &= \sum_{j=0}^n p_{j+1} \mu_{j+1} = R \end{aligned}$$

Logo:

$$G_\epsilon = \frac{1}{PR} \sum_{k=0}^{2n+1} \sum_{l < k} q_l q_k (s_k - s_l) \quad (21)$$

Examinemos agora $\lim_{\epsilon \rightarrow 0} G_\epsilon$.

Na fórmula (21), para um valor fixado de $k = 2j$, com $j = 0, \dots, n$, o somatório interno fica:

$$\begin{aligned} & (q_0 q_k (s_k - s_0) + q_1 q_k (s_k - s_1)) + \dots + (q_{k-2} q_k (s_k - s_{k-2}) + q_{k-1} q_k (s_k - s_{k-1})) \\ &= \sum_{i=1}^j q_{2j} (q_{2i-2} (s_{2j} - s_{2i-2}) + q_{2i-1} (s_{2j} - s_{2i-1})) \end{aligned}$$

Utilizando as expressões (20) e (19), verifica-se que, quando $\epsilon \rightarrow 0$, esta soma tende a:

$$\begin{aligned} & \sum_{i=1}^j \alpha_{j+1} p_{j+1} (\alpha_i p_i (r_j - r_{i-1}) + (1 - \alpha_i) p_i (r_j - r_i)) \\ &= \sum_{i=1}^j \alpha_{j+1} p_{j+1} p_i (r_j - \mu_i) \end{aligned}$$

Portanto essas parcelas contribuem para $\lim_{\epsilon \rightarrow 0} G_\epsilon$, com o valor:

$$\frac{1}{PR} \sum_{j=0}^n \sum_{i=1}^j \alpha_{j+1} p_{j+1} p_i (r_j - \mu_i) \quad (22)$$

Por outro lado, para um valor ímpar de $k = 2j+1$, com $j = 0, \dots, n$, o somatório interno em (21) fica:

$$\begin{aligned} & (q_0 q_k (s_k - s_0) + q_1 q_k (s_k - s_1)) + \dots + (q_{k-3} q_k (s_k - s_{k-3}) + q_{k-2} q_k (s_k - s_{k-2})) + \\ & + q_{k-1} q_k (s_k - s_{k-1}) = \\ & \sum_{i=1}^j q_{2j+1} (q_{2i-2} (s_{2j+1} - s_{2i-2}) + q_{2i-1} (s_{2j+1} - s_{2i-1})) + q_{2j} q_{2j+1} (s_{2j+1} - s_{2j}) \quad (23) \end{aligned}$$

Quando $\epsilon \rightarrow 0$, a última parcela de (23), a saber, $q_{2j} q_{2j+1} (s_{2j+1} - s_{2j})$, tende a:

$$\begin{aligned} & \alpha_{j+1} p_{j+1} (1 - \alpha_{j+1}) p_{j+1} (r_{j+1} - r_j) \\ &= \alpha_{j+1} (1 - \alpha_{j+1}) p_{j+1}^2 (r_{j+1} - r_j) \end{aligned}$$

Então, estas parcelas contribuem para $\lim_{\epsilon \rightarrow 0} G_\epsilon$, com o valor:

$$\frac{1}{PR} \sum_{j=0}^n \alpha_{j+1} (1 - \alpha_{j+1}) p_{j+1}^2 (r_{j+1} - r_j) \quad (24)$$

Finalmente, ainda por (19) e (20), as outras parcelas de (23) tendem a:

$$\begin{aligned} & \sum_{i=1}^j (1 - \alpha_{j+1}) p_{j+1} (\alpha_i p_i (r_{j+1} - r_{i-1}) + (1 - \alpha_i) p_i (r_{j+1} - r_i)) \\ &= \sum_{i=1}^j (1 - \alpha_{j+1}) p_{j+1} p_i (r_{j+1} - \mu_i) \end{aligned}$$

Portanto essas parcelas contribuem para $\lim_{\epsilon \rightarrow 0} G_\epsilon$, com o valor:

$$\frac{1}{PR} \sum_{j=0}^n \sum_{i=1}^j (1 - \alpha_{j+1}) p_{j+1} p_i (r_{j+1} - \mu_i) \quad (25)$$

Em suma: $\lim_{\epsilon \rightarrow 0} G_\epsilon = (22) + (24) + (25)$.

Porém, associando (22) com (25) e utilizando (19), obtém-se:

$$\begin{aligned} & \frac{1}{PR} \sum_{j=0}^n \sum_{i=1}^j p_{j+1} p_i (\alpha_{j+1} (r_j - \mu_i) + (1 - \alpha_{j+1}) (r_{j+1} - \mu_i)) \\ &= \frac{1}{PR} \sum_{j=0}^n \sum_{i=1}^j p_i p_{j+1} (\mu_{j+1} - \mu_i) \\ &= \frac{1}{PR} \sum_{j=1}^{n+1} \sum_{i < j} p_i p_j (\mu_j - \mu_i) = G_I \end{aligned}$$

em virtude de (18).

Por outro lado, utilizando (19) e o fato de que $P\mu = R$, verifica-se que a fórmula (25) é idêntica à (14). A conclusão é que:

$$\lim_{\epsilon \rightarrow 0} G_\epsilon = G_I + D = G_S$$

Resumindo (em termos informais): a cota superior G_S para o índice de Gini corresponde à situação limite em que, em cada classe de renda, as rendas das pessoas se concentram em dois grupos: um com a renda mínima da classe e outro com uma renda "arbitrariamente próxima" da renda mínima da classe seguinte.

12. Exemplos

Para efeito de ilustração, foram tomados os dados de rendimento das pessoas com rendimento, investigados pelo IBGE na Pesquisa Nacional por Amostra de Domicílios - PNAD, para o Brasil e o Estado do Rio de Janeiro, nos anos de 1984, 1985 e 1986. As classes são as fixadas pela própria publicação do IBGE e os rendimentos médios de cada classe são dados da pesquisa. Para efeito de cálculo do índice de Gini, foi atribuída à classe aberta um limite superior igual à diferença entre o dobro da média e o limite inferior da classe. Com os dados constantes das Tabelas 2, 3 e 4, foram calculadas as cotas inferior e superior para o índice de Gini, apresentadas na Tabela 1.

Tabela 1
Cotas Inferior e Superior para o Índice de Gini
Brasil e Rio de Janeiro - 1984, 1985 e 1986

	Brasil-84	Brasil-85	Brasil-86	RJ-84	RJ-85	RJ-86
G_I	0,5756	0,5877	0,5776	0,5608	0,5818	0,5737
G_S	0,5919	0,6039	0,5935	0,5766	0,5972	0,5891

Fonte: Dados das Tabelas 2, 3 e 4.

Tabela 2
Pessoas Ocupadas e Rendimento Médio
Segundo Classes de Rendimento - 1984
Brasil

Classes (em salários mínimos)	Pessoas (ocupadas com rendimento)	Rendimento Médio (em salários mínimos)
0,0 - 0,5	5 822 120	0,2921
0,5 - 1,0	9 541 479	0,7890
1,0 - 2,0	11 932 181	1,3940
2,0 - 3,0	6 030 514	2,3802
3,0 - 5,0	5 491 357	3,7317
5,0 - 10,0	4 135 434	6,7505
10,0 - 20,0	1 731 303	13,3785
20,0 ou mais	821 915	31,2648

Rio de Janeiro

Classes (em salários mínimos)	Pessoas (ocupadas com rendimento)	Rendimento Médio (em salários mínimos)
0,0 - 0,5	471 456	0,2837
0,5 - 1,0	983 913	0,8456
1,0 - 2,0	1 177 709	1,4089
2,0 - 3,0	664 918	2,3733
3,0 - 5,0	653 597	3,7576
5,0 - 10,0	475 572	6,7022
10,0 - 20,0	201 513	13,1712
20,0 ou mais	104 601	29,0327

Fonte: IBGE - PNAD

Tabela 3
Pessoas Ocupadas e Rendimento Médio
Segundo Classes de Rendimento - 1985
Brasil

Classes (em salários mínimos)	Pessoas (ocupadas com rendimento)	Rendimento Médio (em salários mínimos)
0,0 - 0,5	6 491 906	0,2911
0,5 - 1,0	9 688 267	0,8084
1,0 - 2,0	12 825 056	1,4293
2,0 - 3,0	5 945 925	2,4129
3,0 - 5,0	6 231 902	3,8100
5,0 - 10,0	4 608 518	6,9703
10,0 - 20,0	2 006 178	13,8144
20,0 ou mais	925 451	34,6460

Rio de Janeiro

Classes (em salários mínimos)	Pessoas (ocupadas com rendimento)	Rendimento Médio (em salários mínimos)
0,0 - 0,5	518 327	0,2943
0,5 - 1,0	1 123 553	0,8667
1,0 - 2,0	1 202 879	1,4488
2,0 - 3,0	622 362	2,4156
3,0 - 5,0	715 397	3,8470
5,0 - 10,0	534 810	6,9580
10,0 - 20,0	233 370	13,7260
20,0 ou mais	125 993	32,5291

Fonte: IBGE - PNAD

Tabela 4
Pessoas Ocupadas e Rendimento Médio
Segundo Classes de Rendimento - 1986
Brasil

Classes (em salários mínimos)	Pessoas (ocupadas com rendimento)	Rendimento Médio (em salários mínimos)
0,0 - 0,5	4 508 052	0,3408
0,5 - 1,0	9 728 471	0,8632
1,0 - 2,0	12 925 195	1,5087
2,0 - 3,0	6 845 757	2,4776
3,0 - 5,0	7 834 865	3,8918
5,0 - 10,0	5 482 568	7,0373
10,0 - 20,0	2 512 151	14,1132
20,0 ou mais	1 223 570	38,1878

Rio de Janeiro

Classes (em salários mínimos)	Pessoas (ocupadas com rendimento)	Rendimento Médio (em salários mínimos)
0,0 - 0,5	334 357	0,3420
0,5 - 1,0	1 152 212	0,8943
1,0 - 2,0	1 300 134	1,5274
2,0 - 3,0	681 473	2,5012
3,0 - 5,0	843 285	3,9067
5,0 - 10,0	547 299	7,1468
10,0 - 20,0	285 869	14,4838
20,0 ou mais	143 197	35,2935

Fonte: IBGE - PNAD

Bibliografia

[1] **Gastwirth, J.L.** , The estimation of the Lorenz curve and Gini index ; *The Review of Economics and Statistics*, 54, (1972), 306-316.

[2] **Kendall, M.G. & Stuart, A.** , *The advanced theory of Statistics*; London, Charles Griffen and Company, (1963).

[3] **Yntema, D.** , Measures of inequality in the personal distribution of wealth and income ; *Journal of the American Statistical Association*, 28, (1933), 423-433.

Apêndice

Curva de Lorenz, Índice de Gini e Discrepância Máxima para a Distribuição Lognormal

1. Curva de Lorenz

Uma variável aleatória positiva X é lognormal de parâmetros μ e σ , por definição, quando $\log X$ for $N(\mu, \sigma)$, isto é, normal com média μ e desvio-padrão σ . Neste caso, a função de distribuição de probabilidades F de X é nula para $x \leq 0$ e vale, para $x > 0$:

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\log x} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} dt = N_{\mu,\sigma}(\log x) = N_{0,1}\left(\frac{\log x - \mu}{\sigma}\right)$$

onde $N_{\mu,\sigma}$ representa a integral até x da normal $N(\mu, \sigma)$.

É imediato calcular que, nesse caso:

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

$$\mu^* = e^{\mu + \frac{\sigma^2}{2}}$$

$$\Phi(x) = \frac{1}{\mu^*} \int_0^x t f(t) dt = N_{0,1}\left(\frac{\log x - \mu}{\sigma} - \sigma\right) = N_{\mu+\sigma^2,\sigma}(\log x)$$

onde μ^* é a média de X .

Como $u = F(x) = N_{\mu,\sigma}(\log x)$, então:

$$v = \Phi(F^{-1}(u)) = N_{\mu+\sigma^2,\sigma}(N_{\mu,\sigma}^{-1}(u))$$

Uma mudança de variável mostra que $N_{\mu,\sigma}(x) = N_{0,1}\left(\frac{x-\mu}{\sigma}\right)$ e, portanto, $N_{\mu,\sigma}^{-1}(y) = \mu + \sigma N_{0,1}^{-1}(y)$. Segue-se que:

$$\begin{aligned} v &= N_{\mu+\sigma^2,\sigma}(\mu + \sigma N_{0,1}^{-1}(u)) = N_{0,1}\left(\frac{\mu + \sigma N_{0,1}^{-1}(u) - \mu - \sigma^2}{\sigma}\right) \\ &= N_{0,1}(N_{0,1}^{-1}(u) - \sigma) \end{aligned}$$

2. Discrepância máxima

$$\begin{aligned}\delta &= F(\mu^*) - \Phi(\mu^*) = N_{\mu, \sigma}(\mu + \frac{\sigma^2}{2}) - N_{\mu + \sigma^2, \sigma}(\mu + \frac{\sigma^2}{2}) \\ &= N_{0,1}(\frac{\sigma}{2}) - N_{0,1}(-\frac{\sigma}{2}) = 2N_{0,1}(\frac{\sigma}{2}) - 1\end{aligned}$$

Pode-se também escrever $\delta = 2\Psi(\frac{\sigma}{2})$, onde :

$$\Psi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt = N_{0,1}(x) - \frac{1}{2}$$

3. Índice de Gini

Pondo $g(x) = \frac{\log x - \mu}{\sigma}$:

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{g(x)} e^{-\frac{t^2}{2}} dt \quad 1 - F(x) = \frac{1}{\sqrt{2\pi}} \int_{g(x)}^{+\infty} e^{-\frac{t^2}{2}} dt$$

Fixado x , escrevamos $\alpha = g(x)$. Então:

$$F(x)(1 - F(x)) = \frac{1}{2\pi} \int_{-\infty}^{\alpha} e^{-\frac{t^2}{2}} ds \int_{\alpha}^{+\infty} e^{-\frac{t^2}{2}} dt$$

Transforma-se agora esse produto de integrais em uma integral dupla:

$$2\pi F(x)(1 - F(x)) = \int_{-\infty}^{\alpha} \int_{\alpha}^{+\infty} e^{-\frac{s^2+t^2}{2}} ds dt = I$$

Para transformar para coordenadas polares, uma figura conveniente mostrará que:

(i) Se $\alpha > 0$, então:

$$\begin{aligned}I &= \int_{\frac{\pi}{4}}^{\frac{3\pi}{4}} \int_{\alpha \cos \theta}^{\alpha \sec \theta} e^{-\frac{r^2}{2}} r dr d\theta + \int_{\frac{\pi}{2}}^{\pi} \int_{\alpha \sec(\theta - \frac{\pi}{2})}^{+\infty} e^{-\frac{r^2}{2}} r dr d\theta \\ &= \int_{\frac{\pi}{4}}^{\frac{3\pi}{4}} \left(e^{-\frac{\alpha^2 \cos^2 2\theta}{2}} - e^{-\frac{\alpha^2 \sec^2 2\theta}{2}} \right) d\theta + \int_{\frac{\pi}{2}}^{\pi} e^{-\frac{\alpha^2 \csc^2 2\theta}{2}} d\theta \\ &= \int_{\frac{\pi}{4}}^{\frac{3\pi}{4}} e^{-\frac{\alpha^2 \cos^2 2\theta}{2}} d\theta - \int_{\frac{\pi}{4}}^{\frac{3\pi}{4}} e^{-\frac{\alpha^2 \sec^2 2\theta}{2}} d\theta + \int_{\frac{\pi}{2}}^{\pi} e^{-\frac{\alpha^2 \csc^2 2\theta}{2}} d\theta\end{aligned}$$

Fazendo $\gamma = \frac{\pi}{2} - \theta$ na primeira parcela e $\gamma = \theta - \frac{\pi}{2}$ na terceira, vem:

$$\begin{aligned} I &= \int_0^{\frac{\pi}{4}} e^{-\frac{\alpha^2 \sec^2 \gamma}{2}} d\gamma - \int_{\frac{\pi}{4}}^{\frac{\pi}{2}} e^{-\frac{\alpha^2 \sec^2 \theta}{2}} d\theta + \int_0^{\frac{\pi}{2}} e^{-\frac{\alpha^2 \sec^2 \gamma}{2}} d\gamma \\ &= 2 \int_0^{\frac{\pi}{4}} e^{-\frac{\alpha^2 \sec^2 \theta}{2}} d\theta \end{aligned}$$

(ii) Se $\alpha < 0$, então:

$$\begin{aligned} I &= \int_{\frac{\pi}{2}}^{\pi} \int_{-\alpha \sec(\theta-\pi)}^{+\infty} e^{-\frac{r^2}{2}} r dr d\theta + \int_{\pi}^{\frac{5\pi}{4}} \int_{-\alpha \sec(\theta-\pi)}^{-\alpha \operatorname{cosec}(\theta-\pi)} e^{-\frac{r^2}{2}} r dr d\theta \\ &= \int_{\frac{\pi}{2}}^{\pi} e^{-\frac{\alpha^2 \sec^2 \theta}{2}} d\theta + \int_{\pi}^{\frac{5\pi}{4}} \left(e^{-\frac{\alpha^2 \sec^2 \theta}{2}} - e^{-\frac{\alpha^2 \operatorname{cosec}^2 \theta}{2}} \right) d\theta \\ &= \int_{\frac{\pi}{2}}^{\pi} e^{-\frac{\alpha^2 \sec^2 \theta}{2}} d\theta + \int_{\pi}^{\frac{5\pi}{4}} e^{-\frac{\alpha^2 \sec^2 \theta}{2}} d\theta - \int_{\pi}^{\frac{5\pi}{4}} e^{-\frac{\alpha^2 \operatorname{cosec}^2 \theta}{2}} d\theta \end{aligned}$$

Fazendo $\gamma = \pi - \theta$ na primeira parcela, $\theta - \pi$ na segunda, e $\frac{3\pi}{2} - \theta$ na terceira:

$$\begin{aligned} I &= \int_0^{\frac{\pi}{2}} e^{-\frac{\alpha^2 \sec^2 \gamma}{2}} d\gamma + \int_0^{\frac{\pi}{4}} e^{-\frac{\alpha^2 \sec^2 \gamma}{2}} d\gamma - \int_{\frac{\pi}{4}}^{\frac{\pi}{2}} e^{-\frac{\alpha^2 \sec^2 \gamma}{2}} d\gamma \\ &= 2 \int_0^{\frac{\pi}{4}} e^{-\frac{\alpha^2 \sec^2 \theta}{2}} d\theta \end{aligned}$$

(iii) Se $\alpha = 0$, então:

$$I = \int_{-\infty}^0 e^{-\frac{t^2}{2}} dt + \int_0^{+\infty} e^{-\frac{t^2}{2}} dt = \left(\frac{\sqrt{2\pi}}{2} \right)^2 = \frac{\pi}{2}$$

Por outro lado,

$$2 \int_0^{\frac{\pi}{4}} e^{-\frac{\alpha^2 \sec^2 \theta}{2}} d\theta = 2 \int_0^{\frac{\pi}{4}} d\theta = \frac{\pi}{2}$$

A conclusão é que, para qualquer $x > 0$, isto é, seja qual for $g(x)$, se tem:

$$2\pi F(x)(1 - F(x)) = 2 \int_0^{\frac{\pi}{4}} e^{-\frac{g^2(x) \sec^2 \theta}{2}} d\theta$$

Dai segue que:

$$\frac{\Delta}{2} = \int_0^{+\infty} F(x) (1 - F(x)) dx = \frac{1}{\pi} \int_0^{+\infty} \int_0^{\frac{\pi}{4}} e^{-\frac{z^2(x) \sec^2 \theta}{2}} d\theta dx$$

Trocando a ordem de integração:

$$\frac{\Delta}{2} = \frac{1}{\pi} \int_0^{\frac{\pi}{4}} J(\theta) d\theta$$

onde $J(\theta) = \int_0^{+\infty} e^{-\frac{z^2(x) \sec^2 \theta}{2}} dx$.

Lembrando que $g(x) = \frac{\log x - \mu}{\sigma}$ e pondo $z = g(x) \sec \theta - \sigma \cos \theta$, obtém-se:

$$J(\theta) = \sigma \cos \theta e^{\mu + \frac{\sigma^2 \cos^2 \theta}{2}} \int_{+\infty}^{+\infty} e^{-\frac{z^2}{2}} dz = \sigma \sqrt{2\pi} \cos \theta e^{\mu + \frac{\sigma^2 \cos^2 \theta}{2}}$$

Substituindo:

$$\frac{\Delta}{2} = \sqrt{\frac{2}{\pi}} \sigma e^{\mu} \int_0^{\frac{\pi}{4}} \cos \theta e^{\frac{\sigma^2 \cos^2 \theta}{2}} d\theta$$

Fazendo $\sigma \sin \theta = w$, obtém-se $\sigma \cos \theta d\theta = dw$. Como $e^{\frac{\sigma^2 \cos^2 \theta}{2}} = e^{\frac{\sigma^2 - w^2}{2}}$, e levando em conta o valor de μ^* :

$$\begin{aligned} G &= \frac{\Delta}{2\mu^*} = \sqrt{\frac{2}{\pi}} \int_0^{\frac{\sigma}{\sqrt{2}}} e^{-\frac{w^2}{2}} dw \\ &= 2\psi\left(\frac{\sigma}{\sqrt{2}}\right) = 2N_{0,1}\left(\frac{\sigma}{\sqrt{2}}\right) - 1 \end{aligned}$$

NÚMEROS JÁ PUBLICADOS:

01/88 - CRÍTICA DE RAZÕES NO CENSO ECONÔMICO

ren₂ Martins Assunção (ENCE/IBGE)

Rosa₁ de Freitas Castro (DEIND/IBGE)

José Carlos R.C.Pinheiro (ENCE/IBGE)

02/88 - USO DE AMOSTRAGEM EM SIMULAÇÃO DE LEGISLAÇÃO TRIBUTÁRIA

José Carlos da Rocha C.Pinheiro (ENCE/IBGE)

Manuel Martins Filho (DISUL/SERPRO)

03/88 - FORECASTING THE NUMBER OF AIDS CASES IN BRAZIL

Dani Gamerman (IME/UFRJ)

Helio S. Migon (ENCE/IBGE & IME/UFRJ)

04/88 - AVALIAÇÃO DOS EFEITOS DE REDUÇÃO DA FRAÇÃO DE AMOSTRAGEM NO CENSO
DEMOGRÁFICO

José Carlos da Rocha C.Pinheiro (ENCE)

José Matias de Lima (DPE/NME)

01/89 - MIGRAÇÕES ANUAIS RURAL - URBANO - RURAL

Período 70/80)

Kaizô Iwakami Beltrão (ENCE/IBGE)

Helio dos Santos Migon (ENCE/IBGE)

02/89 - CURVA DE LORENZ & ÍNDICE DE GINI PARA DISTRIBUIÇÕES DE RENDA

José Paulo Q.Carneiro - IBGE/ENCE

Jorge Luiz Rangel Costa - CIDE