O tamanho da amostra na amostragem por unidade monetária

Ângelo Henrique Lopes da Silva
Servidor do Tribunal de Contas da União, graduado em Engenharia Mecânica (ITA), Mestre e Doutor em Economia (UnB).

RESUMO

Pouco empregada no âmbito das auditorias no setor público nacional, a Amostragem por Unidade Monetária possui reconhecimento internacional como técnica útil em auditorias financeiras. A literatura contábil é pródiga em apresentar procedimentos dessa técnica, sendo raros, entretanto, os trabalhos que pormenorizam os seus fundamentos. De outro lado, os estatísticos pátrios pouco compreendem a Amostragem por Unidade Monetária, em virtude de não ser uma técnica tradicional de amostragem. Visando preencher essa lacuna multidisciplinar, este artigo busca apresentar a fundamentação estatística que ampara o dimensionamento amostral da técnica. Propõe, com base em documentações contábeis internacionais e em desenvolvimentos matemáticos, um método geral de cálculo de tamanho de amostra. O método proposto também se presta a justificar a formulação tradicionalmente encontrada na literatura contábil.

Palavras-chave: Auditoria financeira. Amostragem por Unidade Monetária. Distribuições de probabilidade.

1. INTRODUÇÃO

Por ter foco na variável monetária, por dar prioridade a valores vultosos na contabilidade e por permitir menores tamanhos de amostra, a Amostragem por Uni-

dade Monetária tem-se destacado nos trabalhos de auditoria. Entretanto, a técnica possui formas de cálculo e conjunto amplo de procedimentos não convencionais, se comparados às demais técnicas de amostragem.

Investigando as raízes estatísticas da Amostragem por Unidade Monetária, esse trabalho busca apresentar, dentre os presentes na literatura, o método mais adequado e geral para dimensionamento da amostra. A compreensão de tal método de dimensionamento, apesar da demonstração envolver conhecimentos estatísticos de distribuição de probabilidade para sua demonstração, exige apenas que se aplique a função de distribuição Gama à obtenção do resultado final do tamanho da amostra, o que pode ser automatizado com facilidade. Alternativamente, constrói-se um tabelamento de tamanhos de amostra que proporciona a obtenção do dimensionamento.

A primeira seção traz as razões da Amostragem por Unidade Monetária e as informações necessárias para consecução do dimensionamento da amostra. Em seguida, submerge-se no tratamento estatístico para o desvendamento das distribuições de probabilidade relacionadas ao tema. Passa-se ao resultado principal do equacionamento final de obtenção de tamanho de amostras e do tabelamento desses tamanhos. Finalmente, comentários finais são registrados.

2. OBJETIVO E INFORMAÇÕES NECESSÁRIAS

A Amostragem por Unidade Monetária (AUM)[1] objetiva verificar se um determinado balanço contábil contém valores equivocados ou fraudados em quantidade significativa estatisticamente. Como toda técnica de amostragem, é possível, observando apenas uma amostra de contas do balanço, opinar sobre o seu conjunto. Para isso, são necessários quatro informações para proceder-se ao uso da técnica: o montante registrado, a materialidade, a taxa esperada de desvios e o nível de confiança.

O montante registrado (M) é o valor total **registrado** no balanço contábil em exame. A materialidade (MA) representa o valor máximo tolerável das diferenças ou desvios entre o verificado na auditoria e o registrado no balanço, correspondendo ao valor limite de aceitação ou rejeição do exame. Se os desvios financeiros inferidos forem maiores que MA, conclui-se pela baixa fidedignidade do valor total do balanço, resultando em sua rejeição.

A título de exemplo, suponha que o valor total do balanço some M=$10.000.000,00 e que se estabeleça que o valor aceitável de desvios nesse balanço seja de, no máximo, MA=$300.000,00. Assim, procura-se, ao se verificar os desvios de uma amostra, testar se os desvios totais da população ultrapassam o limite superior de $300.000,00. Alternativamente, esse valor de aceitação pode ser expresso pela proporção de desvios em relação ao balanço, ou seja, no caso particular, pelo quociente de MA/M=0,03.

Para esse teste, deve-se escolher o **nível de confiança** (β) com o qual se necessita realizar o exame. Por fim, deve-se também estimar o total de **desvios previstos na amostra** (DP), que será a quantidade de desvios encontrada efetivamente na amostra obtida. Importante notar que o DP não deve ser muito próximo de MA – não mais que 50% –, pois a vantagem do teste reside no fato de se poder testar a quantidade de desvios na população por meio de um número menor de desvios na amostra[2].

Acompanhando ainda o mesmo exemplo, o auditor pode estimar, por exemplo, que não se deparará com mais de DP=$25.000,00 em desvios na amostra. Comumente, os desvios da amostra também são expressos como proporção do montante registrado, no caso, de DP/M=0,0025.

Esses são os quatro parâmetros necessários para o cálculo do tamanho da amostra. A materialidade e os desvios previstos na amostra são tomados da experiência do auditor e do histórico de outros trabalhos. Perceba que, quanto maior a proporção MA/M maior a redução do tamanho da amostra, este aumenta pela maior proporção DP/M. É escolhido, tradicionalmente, como nível de confiança, o valor de 95%, podendo, em razão de levantamentos preliminares favoráveis, tais como exames de controles internos, ser ele reduzido. Quanto maior o nível de confiança, maior será o tamanho da amostra.

3. DISTRIBUIÇÕES UTILIZADAS

O dimensionamento exato da AUM baseia-se na distribuição Hipergeométrica, que modelaria perfeitamente amostragens de população finita e sem reposição. No entanto, visando a uma maior automação do cálculo, releva-se um pouco dessa exatidão, fazendo-se uso de outras distribuições de probabilidade equivalentes à Hipergeométrica, permitidas em razão das circunstâncias encontradas no contexto da AUM, em que comumente a população N é grande e a probabilidade de impropriedades espera-se que seja reduzida.

Iniciando-se por uma distribuição Hipergeométrica, seja um balanço com valor de N unidades monetárias[3], sendo K delas incorretas e N-K corretas. Retira-se, a uma só vez, uma amostra de n unidades monetárias desse balanço. A probabilidade de se encontrar x unidades monetárias incorretas nessa amostra é dada por:

$$P(X=x|N,K,n) = \frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}}, \; x + 0, 1, \ldots, k$$

Assim, $\binom{K}{x}$ é a quantidade de possibilidades (combinação) de se obterem X unidades monetárias incorretas na amostra de K unidades monetárias incorretas na população, $\binom{N-K}{n-x}$ é a quantidade de possibilidades de se obterem n-x unidades monetárias corretas na amostra de N-K unidades monetárias corretas na população, e $\binom{N}{n}$ é a quantidade de possibilidades de se obterem n unidades monetárias na amostra de uma população de tamanho N. É fácil ver que a probabilidade desejada será produto e divisão das combinações apresentadas acima.

Um exercício ilustra bem qual o raciocínio da AUM com o uso da distribuição Hipergeométrica (TCU, 2002, p. 74). Suponha que uma população de N=600 unidades monetárias contenha K=18 erros, ou seja, uma proporção p=0,03. Qual a probabilidade de uma amostra de n=100 dessa população ter todos os itens perfeitos, ou seja, x=0?

$$P(X=0) = \frac{\binom{18}{0}\binom{600-18}{100-0}}{\binom{600}{100}} \cong 0{,}05$$

Então, na retirada de uma amostra de uma população com taxa de erro de 3%, o risco de essa amostra conter apenas elementos perfeitos é de 5%. Em busca de nosso objetivo, uma interpretação mais desejável pode ser a de que, com nível de confiança de β = 95% (=100%-5%), pode-se afirmar que essa amostra de unidades monetárias sem erros provenha de uma população com erro máximo de 3%.

O mesmo resultado pode ser alcançado pela distribuição **Binomial**. Quando o tamanho da população é suficientemente grande – o que provavelmente acontece em um balanço contábil –, uma distribuição Hipergeométrica de parâmetros (N,K,n) aproxima-se a uma Binomial de parâmetros (n,p), em que p=K/N (Apêndice A). Usando o exemplo numérico acima, podemos verificar que o resultado pela distribuição Binomial é o mesmo, ou seja, de aproximadamente 5% de probabilidade de que, em uma amostra de tamanho 100 sem erros, provenha de uma população de erro máximo de 3% (=18/600).

$$P(X=0) = \binom{100}{0} \cdot 0{,}03^{0} \cdot (1-0{,}03)^{100} \cong 0{,}05$$

Prosseguindo no caminho já anunciado da equivalência entre distribuições, emprega-se desta vez uma distribuição das mais utilizadas na amostragem, sobretudo em aplicativos de auditoria: a distribuição **Poisson**. Quando a taxa de erros p é pequena, como nos testes substantivos de auditoria, e **n** é suficientemente grande, de forma a que **n·p** (número esperado de erros na população) permaneça fixo, a distribuição Binomial de parâmetros (n,p) aproxima-se de uma Poisson de parâmetro único de valor n·p (Apêndice B). Novamente, utilizando o exemplo particular, encontra-se a mesma probabilidade calculada.

$$P(X = 0) = \frac{(100.0,03)^0\, e^{-(100\,.\,0,03)}}{0!} \cong 0,05$$

Pode-se demonstrar que a distribuição de Poisson apresenta o cálculo de tamanho de amostra mais conservador. Ocorre que o tamanho de amostra da modelagem por Poisson é levemente maior do que o obtido pela Binomial, que, por sua vez, supera levemente o mesmo tamanho pela Hipergeométrica (Stewart, 2013, p.15). Essas majorações de tamanho de amostra não são significativas e ainda garantem inferências mais seguras para AUM.

A distribuição Poisson é mais fácil de se trabalhar para a obtenção do tamanho da amostra, pois apenas utiliza dois parâmetros: n e p. A distribuição Poisson seria o ponto final da sequência de distribuições, não fosse ela uma distribuição discreta. Para superar essa desvantagem, a distribuição empregada no cálculo de dimensionamento da amostra será a distribuição contínua **Gama**. A densidade de distribuição Poisson de parâmetro n·p equivale a uma Gama de parâmetro (α+1,1). Como se precisa da distribuição acumulada para se dimensionar a amostra, verifica-se a necessidade de uso da seguinte relação (Apêndice C).

$$P_{Poisson(n\,.\,p)}(X \leq \alpha) = 1 - P_{Gama\,(\alpha\,+\,1,1)}(X \leq n\,.\,p)$$

4. TAMANHO DA AMOSTRA

Para o dimensionamento adequado da amostra na AUM, será utilizada a distribuição Gama, de variável contínua, justificada pela sequência de equivalências entre as distribuições demonstradas na seção anterior (Hipergeométrica → Binomial → Poisson → Gama). Não se olvide que o elemento amostral não são as contas do balanço *per si*, mas sim as unidades monetárias dessas contas. Portanto, n, o tamanho da amostra, não significa o número da contas a serem selecionadas, mas sim o número de unidades monetárias[4].

Perceba que, no último equacionamento, o α é o limite superior da distribuição acumulada de Poisson e está ele relacionado ao número de erros encontrados na amostra. Em outros termos, pode-se fazer $\alpha = n \cdot p_a$, em que pa é a proporção entre os erros da amostra e o tamanho de amostra n. O p_a corresponde aos desvios previstos na amostra (DP) descritos anteriormente. Já o parâmetro da distribuição de Poisson **n·p** equivale ao produto entre n, o tamanho da amostra, e p_t, a proporção de desvios máximo na população a ser testado, correspondendo este último ao quociente entre a materialidade e o montante registrado, MA/M. Substituindo p_a e p_t na última equação de equivalência, obtém-se

$$P_{Poisson(n\,.\,p_t)}(X \leq n\,.\,p_a) = 1 - P_{Gama\,(n\,.\,p_a\,+\,1,1)}(X \leq n\,.\,p_t)$$

Como o interesse é de que:

$$P_{Poisson(n\,.\,p_t)}(X \leq n\,.\,p_a) \leq 1 - \beta$$

Tem-se que:

$$P_{Gama\,(n\,.\,p_a\,+\,1,1)}(X \leq n\,.\,p_t) = \beta$$

em que a desigualdade é substituída pela igualdade, em razão de a distribuição Gama ser contínua e, assim, ser sempre possível obter n que atenda o último equacionamento. O tamanho da amostra n pode ser obtido pela entrada de três parâmetros: o nível de confiança (β), a proporção máxima de desvios na população (p_t) e a proporção máxima de desvios estimados a serem encontrados na amostra (p_a). O tamanho da amostra **n** é obtido por cálculo interativo, pois essa variável não é separável na formulação acima. No entanto, n pode ser obtido facilmente pelo *Microsoft Excel* por intermédio da função DISTGAMA($n \cdot p_t$;$n \cdot p_a$+1;1;1).

Resgatando os dados de entrada do exemplo da primeira seção, observa-se que p_a=DP/M=0,25%, p_t=MA/M=3% e β=95%. Portanto, n pode ser extraído, sem dificuldades, pela ferramenta "atingir meta" do Excel na fórmula "DISTGAMA($n \cdot p_t$;$n \cdot p_a$+1;1;1)=0,95". O valor obtido será de 119 unidades monetárias a serem amostradas.

Na falta de recursos computacionais para o cálculo da amostra, é salutar que sejam tabelados previamente os tamanhos de amostra n em função de p_a e p_t mais prováveis e dos níveis de confiança mais comuns, como 95 e 90% (Tabelas 1 e 2). Os tamanhos de amostra maiores que 500 unidades monetárias foram omitidos nessas tabelas. Para uso dessas tabelas, a pergunta a ser feita seria: qual o tamanho de amostra que, sendo encontrada proporção de desvios totais de, no máximo, p_a, possibilite que se conclua que, com grau de certeza β, o desvio total na população será de, no máximo, p_t? O valor de tamanho de amostra 119 do exemplo anterior pode ser visualizado na primeira tabela.

Caso as proporções pt e pa não estejam registradas nas entradas das tabelas em exato valor, toma-se, por conservadorismo, o maior tamanho amostral do intervalo em que as proporções se encontram. Como exemplo, suponha p_t=4,10% e p_a=0,75% e β=95%. Entre 103 e 113, por segurança, o tamanho de amostra dever ser de 113 unidades monetárias. Suponha agora p_t=4,25% e p_a=0,8%. Entre 103 e 119, o tamanho conservador da amostra deve ser de 119 unidades monetárias. Suponha ainda p_t=4,00% e p_a=0,8%. Entre 103, 113, 119 e 131, a amostra terá tamanho de 131 unidades. Portanto, caso não se tenha a formulação exata, a solução, por meio da tabela, será sempre a maior amostra entre as possibilidades dos extremos dos intervalos de p_t e p_a.

Intuitivamente, as tabelas fazem sentido. A primeira coluna traz tamanhos de amostra suficientes para se concluir sobre a proporção de desvios na população p_t, quando nenhum desvio é encontrado na amostra (p_a=0). Obviamente, quanto maior for p_t, mais provável de se deparar com desvios na amostra, portanto, não necessitando que amostra seja muito grande. De outro lado, dado um nível de desvios na população p_t, quanto maior for a proporção de desvios na amostra, maior será o tamanho da amostra necessária para demonstrar que a população testada não ultrapassa p_t. Finalmente, se a exigência do grau de confiança é menor, como se ver da primeira para a segunda tabela, os tamanhos de amostra reduzem-se.

A tabela também é consistente com a formulação indicada pelo manual de técnicas de amostragem para auditoria do TCU (2002, p. 73) para o caso de desvios esperados na amostra serem nulos (p_a=0). No referido documento, a formulação adotada é mais simples ($n = C_{0,95\%}/p_t = 3/p_t$ ou $n = C_{0,90\%}/p_t = 2,3 / p_t$), cujos resultados coincidem aos da primeira coluna das tabelas 1 e 2. Os índices de confiabilidade são usados também na extrapolação dos erros na fase inferencial de análise dos desvios após a seleção, não sendo necessários no dimensionamento da amostra, consoante o que buscamos neste trabalho.

Os resultados das tabelas também se coadunam com a formulação encontrada em vários documentos internacionais de auditoria financeira, como o da *European Comission* (2013, p. 117). A formulação do tamanho de amostra que permite antecipação de erros na amostra é:

$$n = \frac{M \, . \, C_{0,\beta}}{MA + DP \, . \, FE}$$

Em que FE é o fator de expansão, baseado no nível de confiança β e tabelado pela literatura (*European Comission*, 2013, p. 117). Para o nosso caso, de β=95%, podemos encontrar que FE=1,6. O índice de confiabilidade $C_{0,\beta}$ é igual a 3. Substituindo na fórmula, obtemos 116, que é um pouco menor do que 119, que encontramos usando a tabela 1. Isso reside no fato, como já comentado, de que o tamanho de amostra calculado pelas distribuições Gama e Poisson são mais conservadoras e, por conseguinte, mais seguras do que outras

Tabela 1
tamanho da amostra para nível de confiança β=95%

pt (%)	pa (%)												
	0	**0,25**	**0,50**	**0,75**	**1,00**	**1,25**	**1,50**	**1,75**	**2,00**	**2,25**	**2,50**	**2,75**	**3,00**
1,00	300	-	-	-	-	-	-	-	-	-	-	-	-
1,25	240	370	-	-	-	-	-	-	-	-	-	-	-
1,50	200	285	439	-	-	-	-	-	-	-	-	-	-
1,75	172	231	330	-	-	-	-	-	-	-	-	-	-
2,00	150	195	262	374	-	-	-	-	-	-	-	-	-
2,25	134	168	217	293	418	-	-	-	-	-	-	-	-
2,50	120	147	185	240	324	462	-	-	-	-	-	-	-
2,75	109	132	161	203	263	355	-	-	-	-	-	-	-
3,00	100	**119**	143	175	220	285	385	-	-	-	-	-	-
3,25	93	108	130	154	189	237	308	416	-	-	-	-	-
3,50	86	99	116	137	165	202	255	330	446	-	-	-	-
3,75	80	92	106	124	146	176	216	272	353	476	-	-	-
4,00	75	85	98	**113**	**131**	156	187	230	289	375	-	-	-
4,25	71	80	90	**103**	**119**	139	165	198	243	306	397	-	-
4,50	67	75	84	95	109	126	147	174	209	257	323	419	-
4,75	64	71	79	89	100	115	132	155	183	220	270	340	442
5,00	60	67	74	83	93	105	120	139	162	192	231	284	357

distribuições. O cálculo da amostra pelas distribuições apresentada neste artigo, em substituição à formulação acima, é mais conservadora, desde que o quociente p_a/p_t seja pequeno e não muito próximo a 0,5.

5. COMENTÁRIOS FINAIS

A abordagem de cálculo apresentada neste artigo é baseada naquela adotada pela AICPA – *American Institute of Certified Public Accountants* (Stewart, 2013) e representa a forma mais geral de dimensionamento amostral da técnica de Amostragem por Unidade Monetária, uma vez que se pode encontrar na literatura outras formulações mais simples, atendendo bem em algumas situações, mas em outras, pelo contrário, tornando o cálculo exageradamente dimensionado.

Mais uma vez, vale ressaltar que se deixam a forma de análise e a seleção amostral na Amostragem por Unidade Monetária, que possui característica ótima de ser uma amostragem que prioriza a materialidade na extração da amostra, para um próximo artigo.

REFERÊNCIAS

BRASIL. *Tribunal de Contas da União*. Técnicas de Amostragem para Auditorias. Brasília: TCU, Secretaria-Geral de Controle Externo, 2002.

CARMELO, Sérgio. *Amostragem em Revisão/auditoria:* uma aplicação prática do método PPS. Disponível em: http://www.paginas.esce.ips.pt/disciplinas/Auditoria/Ficheiros%20diversos/Artigos/Amostragem.pdf. Acesso em: 01.jan.2013.

CASELLA, George; BERGER, Roger L. *Statistical Inference*. Duxbury Press, 2nd ed, 2002.

EUROPEAN COMISSION. *Guidance Note on Sampling Methods for Audit Authorities*. Disponível em: http://www.interact-eu.net/downloads/7847/COCOF_Guidelines_on_Sampling_revised_11_04_2013.pdf. Acesso em: 01.jan.2013.

OLIVEIRA, Paulo Henrique F.C. *Amostragem Básica*: aplicação em auditoria. Editora Ciência Moderna, 2004.

STEWART, Trevor. *Technical Notes on the AICPA Audit Guide: audit sampling*. Disponível em: <http://www.aicpa.org/Publications/AccountingAuditing/KeyTopics/DownloadableDocuments/Sampling_Guide_Technical_Notes.pdf>. Acesso em: 01.jan.2013.

NOTAS

1 A amostragem por unidade monetária é conhecida por *monetary unit sampling* (MUS) ou *dollar unit sampling* (DUS).

Tabela 2
tamanho da amostra para nível de confiança β=90%

pt (%)	pa (%)												
	0	0,25	0,50	0,75	1,00	1,25	1,50	1,75	2,00	2,25	2,50	2,75	3,00
0,75	308	-	-	-	-	-	-	-	-	-	-	-	-
1,00	231	383	-	-	-	-	-	-	-	-	-	-	-
1,25	185	273	458	-	-	-	-	-	-	-	-	-	-
1,50	154	212	315	-	-	-	-	-	-	-	-	-	-
1,75	132	173	239	358	-	-	-	-	-	-	-	-	-
2,00	116	146	192	266	400	-	-	-	-	-	-	-	-
2,25	103	126	160	211	293	442	-	-	-	-	-	-	-
2,50	93	111	137	174	229	320	483	-	-	-	-	-	-
2,75	84	99	120	147	187	248	347	-	-	-	-	-	-
3,00	77	90	106	130	158	201	267	373	-	-	-	-	-
3,25	71	82	95	113	136	169	215	285	400	-	-	-	-
3,50	66	75	87	101	120	145	179	229	304	426	-	-	-
3,75	62	70	79	91	107	127	153	190	242	322	453	-	-
4,00	58	65	73	83	96	112	133	161	200	256	341	479	-
4,25	55	61	68	77	87	101	118	140	170	211	269	359	-
4,50	52	57	63	71	80	91	105	123	147	178	221	283	378
4,75	49	54	59	66	74	84	95	110	130	153	186	231	297
5,00	47	51	56	62	69	77	87	99	115	134	160	194	242

2 Na literatura internacional, o montante registrado é conhecido por population *dollar value*; a materialidade, por *tolerable misstatement* (TR); os desvios previstos na amostra, por *expected population exception rate* (EPER); e o complemento do nível de confiança (1-β), por *acceptable risk of incorrect acceptance* (ARIA).

3 Melhor explicando, se o balanço tem valor total de $ 10.000.000,00, a unidade amostral corresponde a cada unidade monetária, sendo o tamanho da população N=10.000.000.

4 Pela Amostragem por Unidade Monetária, a rubrica contábil na qual está inserida a unidade monetária amostrada será selecionada para exame. No entanto, o elemento amostral continua sendo a unidade monetária.

APÊNDICE A

Quando o tamanho da população N aumenta, uma distribuição Hipergeométrica (N,K,n) tende a uma Binomial (n,p), em que p = K/N

$$P(X=x|N,K,n) = \frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}} = \frac{K!}{x!\,(K-x)!} \cdot \frac{(N-K)!}{(n-x)!\,(N-K-n+x)!} \cdot \frac{n!\,(N-n)!}{N!} =$$

$$= \frac{N!}{x!\,(n-x)!} \cdot \frac{(N-K)!}{(N-K-n+x)!} \cdot \frac{K!}{K-x)!} \cdot \frac{(N-n)!}{N!} =$$

$$= \binom{n}{x} \cdot \frac{(N-K)\,(N-K-1)\ldots(N-K-n+x+1)\,.\,K(K-1)\ldots(K-x+1)}{N(N-1)\ldots(N-n+1)}$$

Dividindo por N no numerador e no denominador, tem-se que:

$$P(x) = \binom{n}{x} \cdot \frac{(1-k/N)\,(1-k/N-1/N)\ldots[1-K/N-(n+x+1)/N]\,.\,K/N\,(K/N-1/N)\ldots[K/N-(x+1)/N]}{(1-1/N)\ldots[1-(n+1)/N]} =$$

Fazendo N muito grande e p = (K/N), tem-se que:

$$P(x) = \binom{n}{x} \cdot \underbrace{(1-p)\,(1-p)\ldots(1-p)}_{(n-x)\text{ vezes}} \cdot \underbrace{p(p)\ldots(p)}_{x\text{ vezes}} =$$

$$P(x) = \binom{n}{x} \cdot p^{x} \cdot (1-p)^{n-x} =$$

APÊNDICE B

A Binomial(n,p) aproxima-se da Poisson(n·p), quando o produto n·p é constante e n é muito grande:

$$P(x) = \binom{n}{x} \cdot p^x \cdot (1 - p)^{n - x} =$$

Fazendo C=np, tem-se que p=C/n e

$$P(x) = \frac{n(n-1) \ldots (n-x+1)}{x!} \cdot (c/n)^x \cdot (1 - c/n)^{n-x} = \frac{n(n-1) \ldots (n-x+1)}{x!} \cdot c^x/n \cdot (1 - c/n)^{n-x}$$

$$= \frac{n(n-1) \ldots (n-x+1)}{x!} \cdot (c/n)^x \cdot (1 - c/n)^n \cdot (1 - c/n)^{-x} = \frac{n}{n}\,\frac{n(n-1)}{n} \ldots \frac{(n-x+1)}{n} \cdot \frac{c^x}{x!} \cdot (1 - c/n)^n \cdot (1 - c/n)^{-x}$$

Ocorre que, sendo **c** e **x** fixos, podendo afirmar que:

$$\lim_{n \to \infty} = \frac{n}{n}\,\frac{n(n-1)}{n} \ldots \frac{(n-x+1)}{n} \cdot (1 - c/n)^{-x} = 1 \qquad \text{e} \qquad \lim_{n \to \infty} = (1 - c/n)^n = e^{-c}$$

Logo, $P(x) = \dfrac{c^x e^{-c}}{x!} = \dfrac{np^x e^{-np}}{x!}$

APÊNDICE C

Por Casella e Berger (2002, p. 130), a função de densidade Gama(α,β) é dada por

$$P(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta} \text{ , em que } 0<x<\infty,\ \alpha>0 \text{ e } \beta>0$$

Portanto, a subfamília de densidade Gama(α,1) é: $P(x) = \dfrac{1}{\Gamma(\alpha)} x^{\alpha-1} e^{-x}$.

A função no denominador pode ser representada por: $\Gamma(\alpha) = (\alpha - 1)!$

Então, seja X uma Gama(α+1,1) e α um inteiro positivo. Assim, integrando por partes seguidamente, obtém-se

$$P_{Gama}(X \leq m) = \frac{1}{\alpha!} \int_0^m x^\alpha e^{-x} dx = -\frac{m^\alpha e^{-m}}{\alpha!} + \frac{1}{(\alpha - 1)!} \int_0^m x^{\alpha-1} e^{-x} dx =$$

$$= -P_{Poisson}(X = \alpha) + \frac{1}{(\alpha-1)!} \int_0^m x^{\alpha-1} e^{-x} dx = \ldots = 1 - \sum_{i=0}^{\alpha} P_{Poisson}(X = i) = 1 - P_{Poisson}(X \leq \alpha),$$

em que a Poisson tem parâmetro m. Fazendo m=n·p, pode-se afirmar que:

$$= -P_{Poisson(n\,.\,p)}(X \leq \alpha) = 1 - P_{Gama(\alpha - 1)}(X \leq n\,.\,p),$$