Universidade Estadual da Paraíba
Centro de Ciências e Tecnologia
Departamento de Estatística

Janeide Alves Barbosa

Distribuição de Poisson

Campina Grande
Dezembro de 2014

Janeide Alves Barbosa

Distribuição de Poisson

Trabalho de Conclusão de Curso apresentado ao curso de Bacharelado em Estatística do Departamento de Estatistica do Centro de Ciências e Tecnologia da Universidade Estadual da Paraíba em cumprimento às exigências legais para obtenção do título de Bacharel em Estatística.

Orientadora: Divanilda Maia Esteves

Campina Grande
Dezembro de 2014

B238d Barbosa, Janeide Alves.
Distribuição de Poisson [manuscrito] / Janeide Alves Barbosa. - 2014.
31 p. nao

Digitado.
Trabalho de Conclusão de Curso (Graduação em Estatística) - Universidade Estadual da Paraíba, Centro de Ciências e Tecnologia, 2014.
"Orientação: Profa. Dra. Divanilda Maia Esteves, Departamento de Estatística".

1. Distribuição de Poisson. 2. Probabilidade. 3. Distribuição binomial. I. Título.

21. ed. CDD 519.2

Janeide Alves Barbosa

Distribuição de Poisson

Trabalho de Conclusão de Curso apresentado ao curso de Bacharelado em Estatística do Departamento de Estatística do Centro de Ciências e Tecnologia da Universidade Estadual da Paraíba em cumprimento as exigências legais para obtenção do título de Bacharel em Estatística.

Aprovado em 18/12/2014

Banca Examinadora:

Divanilda M Esteves

Prof. Divanilda Maia Esteves
Universidade Estadual da Paraíba-
DE/CCT
Orientadora

Gustavo H. E.

Prof. Gustavo Henrique Esteves
Universidaden Estadual da Paraíba-
DE/CCT
Examinador

Ricardo Alves de Olinda

Prof. Ricardo Alves de Olinda
Universidade Estadual da Paraíba-
DE/CCT
Examinador

Dedicatória

Dedico este trabalho a todos que acreditaram que eu seria capaz de chegar até aqui e seguir a diante.

Agradecimento

Agradeço primeiramento ao nosso Criador Jeová Deus por me proporcionar a realização desse sonho.

Aos meus pais, irmãos, avós (em memoria), tios e primos, enfim a toda minha família e aos professores.

A minha orientadora,Divanilda Maia Esteves, pela paciência, carinho e compreensão.

Não tenho palavras para expressar os agradecimentos a todos aqueles que passaram e estão em minha vida, citar nomes eu poderia esquecer algumas delas e também não caberia nestas linhas.

Resumo

Este trabalho tem por objetivo desenvolver os principais aspectos da Distribuição de Probablidade de Poisson, introduzida por Siméon Denis Poisson em 1837 no livro intitulado: *Recherches sur la probabilité de jugements en matière criminelle at en matière civile*. A distribuição Poisson pode ser aplicada em diversas áreas, podendo ser usada também como uma aproximação para a variável aleatória binomial. Por fim, será utilizado um conjunto de dados do software R para verificar se tais dados se adequam ao modelo através do teste χ^2(lê-se qui-quadrado). Para simulação de conjunto de dados, também utiliza-se o programa R. Na primeira aplicação observou-se que os dados não se adequaram ao modelo proposto, para os dados simulados confirma-se a teoria que a distribuição de Poisson se aproxima da distribuição binomial. Após a aplicação dos dados simulados verificou-se que a distribuição binomial se aproxima da Poisson como visto teoricamente. A partir dos resultados obtidos sobre o conjunto de dados do R, pode se observar que os mesmos não se ajustaram ao modelo proposto.

Palavras-chave: Distribuição Poisson; Probabilidades; Aplicação.

Abstract

This study aims to discuss the Poisson Probability Distribution, Siméon Denis Poisson introduced in 1837 the book entitled:*Recherches sur la probabilité de jugements en matière criminelle at en matière civile*. The Poisson distribution can be applied in many areas and can also be used as an approximation to the binomial random variable. Finally, a software data set R will be used to verify whether such data fits model through the test χ^2 (pronounced chi-square). For simulation of the data set, we also use the program R. In the first application noted that the data did not fit the proposed model for the second application with the simulated data confirm the theory that the Poisson distribution approaches the binomial distribution. After application of the simulated data was found that the binomial distribution approaches the theoretically Poisson seen. From the results obtained on the R data set, it can be observed that they did not fit the proposed model.

Keywords: Poisson Distribution; Probability; Applications.

Sumário

1 Introdução

De acordo com Ross (2010), a distribuição de probabilidade de Poison foi introduzida por Siméon Denis Poisson em um livro que escreveu a respeito da aplicação da teoria da probabilidade a processos de julgamentos criminais e similares. O título do livro, publicado em 1837, era *Recherches seve la probabilité de jugements en matière criminelle et en matière civile* (Investigação sobre a probabilidade de vereditos em materias criminal e civil).

Segundo Mood e Graybill(1963), a Distribuição de Poisson fornece um modelo realístico para vários fenômenos aleatórios. Como os valores de uma variável aleatória de Poisson são inteiros não-negativos. Por exemplo, contar o número de acidentes fatais no tráfego por semana em um determinado estado, o número de emissões de partículas radioativas por hora, o número de chamadas telefônicas por hora, o número de meteóritos que colidem com um satélite em teste durante uma órbita simples, o número de defeitos por unidades de algum material, o número de falhas por comprimentos.

A distribuição Poisson aparece também em situações em que “eventos" ocorrem em certos instantes (aleatórios) de tempo. Exemplos, um terremoto, entrada de pessoas em determinado estabelecimento (bancos, hospitais, postos de combustível, e assim por diante). A variável aleatória de Poisson encontra uma extensa faixa de aplicações em diversas áreas porque pode ser usada como uma aproximação para a variável aleatória binomial com parâmetros (n, p) no caso particular de n grande e p suficientememte pequeno para que np tenha tamanho moderado. Para que os eventos que ocorrem tenham aproximadamente uma distribuição de Poisson, não é necessário que todos os eventos tenham a mesma probabilidade de ocorrência, mas que todas as probabilidades sejam suficientemente pequenas.

Assim como no caso determinístico, alguns modelos matemáticos relativamente simples parecem ser capazes de descrever uma classe bastante grande de fenômenos (MEYER, 1983). Na prática deve-se ter grande cuidado para evitar usar erroneamente a distribuição de Poisson em dados de contagem, por exemplo, em estudar a distribuição de larvas de insetos sobre alguma área de cultivo, o modelo de Poisson é apto a ser inválido desde que insetos pôem ovos em aglomerados implicando que as larvas são susceptíveis de ser encontradas em grupos, que é incompatível com a hipótese de indepedência de contagens em pequenas subáreas (MOOD e GRAYBILL, 1963).

Este trabalho tem por objetivo fazer uma revisão teórica dos principais aspectos da distribuição de Poisson. Depois como aplicação, será checado se o modelo probabilístico de Poisson explica o número de reclamações sobre os médicos de um serviço de emergência, verificando-se a adequação do modelo através do teste qui-quadrado. Os dados são do pacote *Datasets* do R, sob o nome *Faraway* e dizem respeito a queixas sobre médicos de emergências. Outra aplicação será feita, mas desta vez com dados simulados. Utilizamos também o *software* R para gerar amostras da distribuição $binomial(n, p)$ e em seguida, também usando o teste qui-quadrado de aderência, mostrar-se que a distribuição de Poisson se aproxima adequadamente a distribuição binomial, o que é bem conhecido teoricamente.

2 Fundamentação Teórica

Frequentemente quando um experimento é realizado, há interesse em alguma função do resultado e não no resultado em si. Essas funções são conhecidas como variáveis aleatórias. Como o valor da variável aleatória é determinado pelo resultado do experimento, podemos atribuir probabilidades aos possíveis valores da variável aleatória.

As variáveis aleatórias podem ser classificadas em dois tipos:

(i) discretas - quando os possíveis valores formam um conjunto finito enumerável;

(ii) contínuas - cujos possíveis valores pertecem a um intervalo de números reais e que resultam de uma mensuração.

Neste trabalho o interesse reside no modelo probabilístico de Poisson, classificado como variável aletória discreta, introduzida por Siméon Poisson em 1837. Com base em revisões teóricas, será abordada algumas características desta variavél.

2.1 Distribuição de Poisson

Uma variável aleatória X que pode assumir qualquer um dos valores i= 0, 1, 2, 3..., é chamada variável aleatória de Poisson com parâmetro λ se, para $\lambda > 0$,

$$p(x) = P\{X = x\} = e^{-\lambda}\frac{\lambda^x}{x!} \tag{2.1}$$

A Equação 2.1 define uma função de probabilidade, já que

$$\sum_{x=0}^{\infty} P(x) = e^{-\lambda}\sum_{x=0}^{\infty}\frac{\lambda^x}{x!} = e^{-\lambda}e^{\lambda} = 1.$$

Tanto o valor esperado quanto a variância da variável aleatória de Poisson são iguais ao parâmetro λ . Por definição se X é uma váriável discreta,

$$E(X) = \sum_{x\in s} xP(X = x).$$

Neste caso,

$$
\begin{aligned}
E[X] &= \sum_{x=0}^{\infty} \frac{xe^{-\lambda}\lambda^x}{x!} \\
&= e^{-\lambda} \sum_{x=1}^{\infty} \frac{\lambda}{(x-1)}.
\end{aligned}
$$

fazendo a mudança de variaveis, $j = x - 1$, segue

$$
\begin{aligned}
E[X] &= e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^{x+1}}{j!} \\
&= \lambda e{-}\lambda \sum_{x=0}^{\infty} \frac{\lambda^j}{j!} \\
&= \lambda e^{-\lambda} e^{\lambda} = \lambda.
\end{aligned}
$$

já que

$$
\sum_{j=0}^{\infty} \frac{\lambda^j}{j!} = e^{\lambda}.
$$

Para determinar a variância, primeiro é necessário calcular $E[X^2]$.

Sabe-se que se X é uma variável aleatória discreta assumindo-se valores num conjunto S, então

$$
E(X^2) = \sum_{x \in S} x^2 P(X = x).
$$

logo, se $X \sim Poisson(\lambda)$

$$
\begin{aligned}
E[X^2] &= \sum_{x=0}^{\infty} \frac{x^2 e^{-\lambda}\lambda^x}{x!} \\
&= \lambda \sum_{x=1}^{\infty} \frac{xe^{-\lambda}\lambda^{(x-1)}}{(x-1)!}.
\end{aligned}
$$

Fazendo a transformação $j = x - 1$, tem-se

$$\begin{aligned} E[X^2] &= \lambda \sum_{j=0}^{\infty} \frac{(j+1)e^{-\lambda}\lambda^j}{j!} \\ &= \lambda \left[\sum_{j=0}^{\infty} \frac{je^{-\lambda}\lambda^j}{j!} + \sum_{j=0}^{\infty} \frac{e^{-\lambda}\lambda^j}{j!} \right] \\ &= \lambda(\lambda+1). \end{aligned}$$

Como $E[X] = \lambda$, então

$$\begin{aligned} Var(X) &= E[X^2] - (E[X])^2 \\ &= \lambda(\lambda+1) - \lambda^2 \\ &= \lambda^2 + \lambda - \lambda^2 \\ &= \lambda. \end{aligned}$$

Com isso, o valor esperado e a variância de uma variável aleatória de Poisson são iguais ao seu parâmetro λ.
Outra característica importante das variáveis aleatórias discretas é a Função Geratriz de Momentos (F.G.M), que é definida para todos os valores reais de t, como

$$\begin{aligned} M(t) &= E[e^{tX}] \\ &= \sum_x e^{tx} p(x). \end{aligned}$$

Chama-se $M(t)$ de função geratriz de momentos porque todos os momentos de X podem ser obtidos com o cálculo sucessivo da derivada de $M(t)$ e então com sua avaliação em $t = 0$.Vejamos,

$$M'(t) = \frac{d}{dt} E[e^{tX}] = E\left[\frac{d}{dt}(e^{tX})\right] = E[Xe^{tX}] \tag{2.2}$$

onde adotamos como legítima a hipótese da troca de ordem do cáculo da derivada e da esperança. Isto é, considera-se que

$$\frac{d}{dt}\left[\sum_x e^{tx} p(x)\right] = \sum_x \frac{d}{dt}[e^{tx} p(x)].$$

Portanto, da Equação (2.2), aplicando-se em $t = 0$, obtêm-se

$$M'(0) = E[X].$$

Similarmente vamos a segunda derivada

$$\begin{aligned} M''(t) &= \frac{d}{dt} M'(t) \\ &= \frac{d}{dt} E[Xe^{tX}] \\ &= E\left[\frac{d}{dt}(Xe^{tX})\right] \\ &= E[X^2 e^{tX}]. \end{aligned}$$

Assim,

$$M''(0) = E[X^2].$$

De modo geral, a *n-ésima* derivada de $M(t)$ é dada por

$$M^n(t) = E[X^n e^{tX}] n \geq 1,$$

implicando-se que

$$M^n(0) = E[X^n] n \geq 1.$$

Se X é uma variável aleatória de Poisson com parâmetro λ, então

$$\begin{aligned} M(t) &= E[e^{tX}] \\ &= \sum_{n=0}^{\infty} \frac{e^{tn} e^{-\lambda} \lambda^n}{n!} \\ &= e^{\lambda} \sum_{n=0}^{\infty} \frac{(\lambda e^t)^n}{n!} \\ &= e^{-\lambda} e^{\lambda^{e^t}} \\ &= exp\{\lambda(e^t - 1)\}. \end{aligned}$$

Calculando a primeira derivada, obtemos:

$$M'(t) = \lambda e^t exp\{\lambda(e^t - 1)\},$$

e a segunda derivada é

$$M''(t) = (\lambda e^t)^2 exp\{\lambda(e^t - 1)\} + \lambda e^t exp\{\lambda(e^t - 1)\},$$

assim,

$$E[X] = M'(0) = \lambda.$$

e

$$\begin{aligned} E[X^2] &= M''(0) \\ &= \lambda^2 + \lambda. \end{aligned}$$

De onde segue

$$\begin{aligned} Var(X) &= E[X^2] - (E[X])^2 \\ &= \lambda^2 + \lambda - \lambda \\ &= \lambda. \end{aligned}$$

Portanto, temos outra metodologia para mostrar que a média e a variância de uma variável aleatória de Poisson são iguais a λ.

2.2 Estimador de Máxima Verossimilhança de λ para o Modelo de Poisson

Segundo Bussab e Morettin (2005), o princípio da verossimilhança afirma que deve-se escolher aquele valor do parâmetro desconhecido que maximiza a probabilidade de obter a amostra particular observada, ou seja, o valor que torna aquela amostra a “mais provável". Esse princípio foi enunciado por Ronald Aylmer Fisher pela primeira vez em 1912 e, em 1922, deu-lhe forma mais completa, introduzindo a expressão *likelihood* (verosimilhança).

Em outras palavras, o método de máxima verossimilhança consiste em estimar o parâmetro λ por uma amostra aleatória $X_1, ..., X_n$ independente e igualmente distribuída de X.

Como X segue uma distribuição de Poisson com parâmetro λ, sua distribuição de probabilidade é

$$f(x, \lambda) = \frac{\lambda^x e^{-\lambda}}{x!}. \tag{2.3}$$

Assim, a função de verossimilhança para uma amostra X_1,...,X_n é dada por

$$L(\lambda; x_1, ..., x_n) = \prod_{i=1}^{n} \left(\frac{\lambda^{x_i} e^{-\lambda}}{x_i!} \right) = \frac{e^{-n\lambda} \lambda^{\sum_{i=1}^{n} x_i}}{\prod x_i!}.$$

Aplicando-se o logaritmo natural da função verossimilhança, desta forma tem-se que

$$\begin{aligned}
\ln L(\lambda;x) &= \ln\left[\frac{e^{-n\lambda}\lambda^{\sum_{i=1}^{n}x_i}}{\prod_{i=1}^{n}x_i!}\right]\\
&= \ln e^{-n\lambda} + \ln\lambda^{\sum_{i=1}^{n}x_i} - \ln\prod_{i=1}^{n}x_i!\\
&= -n\lambda + \sum_{i=1}^{n}x_i\ln\lambda - \ln\prod_{i=1}^{n}x_i!.
\end{aligned}$$

O candidato a ponto de máximo da função logaritma de verossimilhanca, é aquele cuja primeira derivada em relação a λ for igual a zero, ou seja,

$$\begin{aligned}
\frac{\partial \ln L(\lambda,x)}{\partial\lambda} &= \frac{1}{\hat{\lambda}}\sum_{i=1}^{n}x_i - n = 0\\
\hat{\lambda} &= \frac{\sum_{i=1}^{n}x_i}{n}\\
\hat{\lambda} &= \bar{X}.
\end{aligned}$$

Para garantir que este é o EMV, deve-se ter a segunda derivada em relação a λ negativa o que de fato acontece, pois

$$\begin{aligned}
\frac{\partial^2 \ln L(\lambda,x)}{\partial\lambda^2} &= \frac{1}{\lambda}\sum_{i=1}^{n}x_i - n = 0\\
&= \frac{\sum_{i=1}^{n}x_i'\lambda - \sum_{i=1}^{n}x_1\lambda'}{\lambda 2} - n'\\
&= -\frac{1}{\lambda^2}\sum_{i=1}^{n}x_i < 0.
\end{aligned}$$

Daí, o estimador de máxima verossimilhança de λ é $\hat{\lambda} = \bar{X}$.

2.2.1 Aproximação da Binomial à Poisson

A variável aleatória de Poisson encontra uma vasta área de aplicação tendo em vista que pode ser usada como uma aproximação para a variável aleatória binomial com parâmetros (n, p) no caso particular de n grande e p pequeno para que np tenha tamanho moderado.

Defini-se a função de densidade binomial com parâmetros n e p , como

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k},$$

para $k = 0, 1, ..., n$.

E o parâmetro n aproxima-se para infinito e p aproxima-se para 0 de modo que $np = \lambda$, isto é,

$$\begin{aligned}
\binom{n}{k} p^k (1-p)^{n-k} &= \frac{n!}{(n-k)!k!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \\
&= \frac{\lambda^k}{k!} \frac{n!}{(n-k)!} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-k} \\
&\to \frac{e^{-\lambda}\lambda^k}{k!},
\end{aligned}$$

uma vez que

$$\left(1 - \frac{\lambda}{n}\right)^{-k} \to 1$$

e

$$\left(1 - \frac{\lambda}{n}\right)^{n} \to e^{-\lambda}$$

com $n \to \infty$. Assim, para n grande e p pequeno a probabilidade binomial $\binom{n}{k} p^k (1-p)^{n-k}$ pode ser aproximada pela probabilidade Poisson $\frac{e^{-np}(np)^k}{k!}$.

2.3 Processo de Poisson

Suponha-se que "eventos" ocorram em instantes aleatórios de tempo e que $N(t)$ represente o número de eventos ocorridos no intervalo $[0, t]$. O conjunto de variáveis aleatórias $\{N(t), t > 0\}$ é chamado de processo de Poisson com taxa λ, $\lambda > 0$, se

(i) $N(0) = 0$, ou seja, o processo começa no instante 0.

(ii) Os números de eventos ocorridos em intervalo de tempo disjuntos forem independentes. Quer dizer que, o número de eventos ocorridos até o tempo t [isto é, N(t)] é independente do número de eventos ocorridos entre t e $t + s$ [isto é, N(t+s)-N(t)].

(iii) A distribuição do número de eventos ocorridos em certo intervalo de tempo depende somente da extensão do intervalo e não de sua localização. A hipótese de incrementos estacionários, diz que a distribuição de probabilidade de $N(t+s) - N(t)$ é a mesma para todos os valores de t.

(iv) $P\{N(h) = 1\} = \lambda h + o(h)$

(v) $P\{N(h) \geqslant 2\} = o(h)$

Uma função é chamada de $o(h)$ se

$$\lim_{h \to 0} \frac{f(h)}{h} = 0.$$

Ou seja, para valores de $h, f(h)$ é pequeno mesmo em relação a h. Conforme visto anteriormente $N(t)$ possui distribuição de Poisson com média λt. Agora realiza-se este resultado por outro método.

Lema1.1

Para uma variável aleatória de Poisson com taxa λ,

$$P\{N(t) = 0\} = e^{-\lambda t}.$$

Demonstração: Seja $P_0(t) = P\{N(t) = 0\}$. Deduzindo uma equação diferencial para $P_0(t)$:

$$\begin{aligned} P_0(t+h) &= P\{N(t+h) = 0\} \\ &= P\{N(t) = 0, N(t+h) - N(t) = 0\} \\ &= P\{N(t) = 0\}P\{N(t+h) - N(t) = 0\} \\ &= P_0(t)[1 - \lambda h + o(h)], \end{aligned}$$

onde as duas últimas equações resultam da condição que, os números de eventos ocorrem em intervalos de tempo disjuntos e independentes, porém os fatos são que $P\{N(h) = 1\} = \lambda h + 0(h)$ e $P\{N(h) \geq 2\} = o(h)$ implicando em $P\{N(h) = 0\} = 1 - \lambda h + o(h)$. Portanto,

$$\tfrac{P_0(t+h) - P_0(t)}{h} = -\lambda P_0(t) + \tfrac{o(h)}{h}.$$

Agora, fazendo-se $h \to o$, obtem-se

$$P_0'(t) = -\lambda P_0(t).$$

Ou, equivalentemente,

$$\frac{P_0'(t)}{P_0(t)} = -\lambda.$$

resultando por integração, em

$$logP_0(t) = -\lambda t + c$$
$$\text{ou}$$
$$P_0(0) = ke^{-\lambda t}.$$

Como $P_0(0) = P\{N(0) = 0\} = 1$, obtem-se $P_0(t) = e^{-\lambda t}$.

Outra característica importante é a relação entrea distribuição de Poisson e a distribuição Exponencial. Na distribuição Poisson estima se a quantidade de eventos num intervalo. Na distribuição Exponencial estima se o intervalo de tempo ou espaço para ocorrência de um evento, dependendo apenas da suposição de que o evento ocorra seguindo o processo de Poisson.

Seja X uma variável aleatória de Poisson com parâmetro λ. Suponha que X representa o número de ocorrências de um evento, num certo intervalo de tempo. Variando-se o intervalo de interesse, pode-se considerar que o número de ocorrências ainda é Poisson, mas com o parâmetro ajustado proporcionalmente. Assim, tem-se a variável $X_t \sim Poisson(\lambda t)$.

Suponha-se que acabamos de ter uma ocorrência e seja T a variável aleatória que indica o tempo até a próxima ocorrência. Tem-se, para $t > 0$,

$$F_T(t) = P(T \leq t) = 1 - P(T > t) = 1 - P(X_t = 0) = 1 - e^{\lambda t}$$

e, assim, concluí-se que T é Exponencial com parâmetro λ.

2.4 Simulando uma variável aleatória de Poisson

Segundo Ross (2010), todos os métodos para simular variáveis aleatorias a partir de distribuições contínuas possuem análogos no caso discreto. Assim, se quiser-se gerar valores de uma variável aleatória discreta X com f.d.p dada por

$$P(X = x_j) = P_j,$$

em que

$$j = 0, 1, ... \qquad \sum_j P_j = 1.$$

Utilizando-se o método da transformada inversa esta geração pode ser feita supondo que U seja uniformemente destribuída entre (0,1). Então, faz se

$$X = \begin{cases} x_0 & se \quad U \leq P_0 \\ x_1 & se \quad P_0 < U \leq P_0 + P_1 \\ \cdots & \\ \cdots & \\ \cdots & \\ x_j & se \quad \sum_{i=0}^{j-1} P_i < U \leq \sum_i^j = P_i. \end{cases}$$

Como

$$P\{X = x_j\} = P\left\{\sum_{i=0}^{j-1} P_i < U \leq \sum_{i=0}^{j} P_i\right\} = P_j$$

,

tem-se que X é a distribuição desejada.

Para simular uma variável aleatória de Poisson com média λ, gera-se variáveis aleatórias $U_1, U_2, ...$ independentes e uniformes no intervalo (0,1) e pare quando

$$N = m\,in\left\{n : \prod_{i=1}^{n} U_i < e^{-\lambda}\right\}$$

Gerando-se números aleatórios continuamente até que seu produto seja menor que $e^{-\lambda}$, então o número necessário, menos 1 ($X = N - 1$), é Poisson com média λ.

Observa-se

$$X + 1 = m\,in\left\{n : \prod_{i=1}^{n} U_i < e^{-\lambda}\right\}$$

é aquivalente a

$$X = m\,ax\left\{n : \prod_{i=1}^{n} U_i \geq e^{-\lambda}\right\},$$

onde

$$\prod_{i=1}^{n} U_{i=1}$$

usando-se logaritmos

$$X = m\,ax\left\{n : \sum_{i=1}^{n} logU_i \geq -\lambda\right\}$$

ou

$$X = m\,ax\left\{n : \sum_{i=1}^{n} -logU_i \leq \lambda\right\}$$

No entanto, $-logU_i$ é exponencial com taxa 1, e com isso X pode ser visto como o número máximo de exponecial com taxa 1 que podem ser somadas e ainda assim o resultado ser menor que λ.

2.4.1 Gerando dados no software R

Segundo Peternelli(2011), o *software* R é uma importante ferramenta na análise e manipulação de dados estatísticos, desde o desenvolvimento e aplicação de métodos até a interpretação de resultados. Além de ser gratuito pode ser obtido em http://www.R-project.org, tem código-fonte aberto, podendo ser modificado ou implementado com novos procedimentos e funções desenvolvidas pelo usuários a qualquer momento.

O R gera número aleatórios de duas maneiras, por um conjunto pre-estabelecido de valores, ou quantos forem necessários de uma distribuição de interesse. Neste caso vamos gera-se números aleatórios de Poisson. Basta usar a codificação “r”, seguido do código da distribuição “pois” e seu parâmetro.

Por exemplo para gerar 20 números aleatórios com média 0,2, usa-se

$> rpois(20, 0.2)$

2.5 Teste de aderência de Qui-quadrado

Como já mencionado, o objetivo deste trabalho é verificar se o conjunto de dados decorrentes de número de reclamações sobre os médicos de emergência, parece seguir uma distribuição de Poisson. Os dados foram observados com 44 médicos. No teste de aderência X^2 têm-se uma população P e verifica-se se ela segue uma distribuição especificada P_0, isto é, queremos testar as hipótese $H_0 : P = P_0$. Também podemos testar essa hipótese, empregando testes sobre os parâmetros média e variância. O teste comparará o número de casos ocorridos em caselas especificadas, com o número esperado de casos nelas, quando a hipótese H_0 for verdadeira. Considera-se classes onde as quais a váriável X, característica da população, pode ser classificada, qualitativa ou quantitativa.

A estatística usada será

$$\chi^2 = \sum_{i=0}^{k} \frac{(o_i - e_i)^2}{e_i},$$

Em que:

k-categoria de provas ou classes;

o_i-frequência observada da i-ésima categoria;

e_i-frequência esperada da i-ésima categoria.

O problema pode ser caracterizado como uma amostra $X_1, ..., X_n$ da variável aleatória X que caracteriza a população P e testa-se a hipótese

$$H_0 : P = P_0,$$

onde P_0 tem uma distribuição de probabilidade especificada. A variável X de interesse da população é categorizada em classes $A_1, A_2, ..., A_s$ e têm-se as probabilidades $p_1 = P(X = i), i \geqslant 0$, $i = 1, 2, ...,$. Visto que χ^2 segue uma distribuição assintótica de χ^2_{k-p-1}, constroí-se a região crítica do teste, ou seja, região de rejeição de H_0.

Decisão: Rejeita-se a hipótese H_0, se o valor calculado da estatística de teste $\chi^2_{calc} > \chi^2_{\alpha,k-p-1}$, em que p representa o número de parâmetros da distribuição estimados pela estatística amostral.

3 Aplicações

Aplica-se nesta seção os métodos teóricos apresentados anteriormente para um particular conjunto de do software R ,dados que conta o número de reclamações sobre médicos em um determinado serviço de emergência de um hospital.

Supondo-se que o conjunto de dados seja explicado pelo modelo probabilístisco de Poisson. Para verificar esta suposição nos foi conveniente aplicar o teste de χ^2(Qui-Quadrado). Tendo em vista que não se conhece o parâmetro λ associado à distribuição a ser testada, usa-se a teoria dos Estimadores de Máxima Verossimilhança (E.M.V.) para obter seu estimador $\hat{\lambda} = \bar{X}$ para enfim, aplicar o teste.

3.1 Aplicação 1

Através do banco de dados do software R, foram observados o número de reclamações sobre 44 médicos de um determinado serviço de emergência. Observe a Tabela 1.

Tabela 1: Frequência do número de reclamações por médicos:

Números de reclamações(X)	Números de médicos(O_i) com (X)reclamações
0	1
1	12
2	12
3	5
4	1
5	4
6	2
7	2
8	2
9	2
10	1
11	1
Total	44

A estimativa do número médio de reclamações para cada médico é a média amostral, dada da seguinte maneira : $\bar{x} = \frac{\sum o_i.x_i}{\sum o_i} = \frac{(1*0+12*1+...+1*11)}{44} = 3,34$. A partir do parâmetro $\hat{\lambda} = 3,34$, calcula-se a probabilidade p_i, correspondente ao número de reclamações. Assumindo os valores 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 e 12, para isso utiliza-se a Equação (1.1).

$$p_0 = P\{X = 0\} = \frac{e^{-3,34}(3,34)^0}{0!} = 0,0354$$

$$p_1 = P\{X = 1\} = \frac{e^{-3,34}(3,34)^1}{1!} = 0,1184$$

$$p_2 = P\{X = 2\} = \frac{e^{-3,34}(3,34)^2}{2!} = 0,1977$$

$$p_3 = P\{X = 3\} = \frac{e^{-3,34}(3,34)^3}{3!} = 0,2201$$

$$p_4 = P\{X = 4\} = \frac{e^{-3,34}(3,34)^4}{4!} = 0,1838$$

$$p_5 = P\{X = 5\} = \frac{e^{-3,34}(3,34)^5}{5!} = 0,1227$$

$$p_6 = P\{X = 6\} = \frac{e^{-3,34}(3,34)^6}{6!} = 0,0683$$

$$p_7 = P\{X = 7\} = \frac{e^{-3,34}(3,34)^7}{7!} = 0,0326$$

$$p_8 = P\{X = 8\} = \frac{e^{-3,34}(3,34)^8}{8!} = 0,0136$$

$$p_9 = P\{X = 9\} = \frac{e^{-3,34}(3,34)^9}{9!} = 0,0051$$

$$p_{10} = P\{X = 10\} = \frac{e^{-3,34}(3,34)^{10}}{10!} = 0,0017$$

$$p_{11} = P\{X = 11\} = \frac{e^{-3,34}(3,34)^{11}}{11!} = 0,0005$$

$$p_{12} = P\{X \geq 12\} = 1 - (p_0 + ... + p_{11}) = 1 - 0.9999 = 0,0001$$

Agora calcula-se as frequências esperadas. Isto é, $ei = np_i$.
Tabela 2: Frequências esperadas do número de reclamações para cada médico.

X	o_i	p_i	$e_i = np_i$
0	1	0,0354	1,5592
1	12	0,1184	5,2078
2	12	0,1977	8,6971
3	5	0,2201	9,6827
4	1	0,1838	8,0851
5	4	0,1227	5,4008
6	2	0,0683	3,0065
7	2	0,0326	1,4345
8	2	0,0136	0,5989
9	2	0,0051	0,2223
10	1	0,0017	0,0742
11	1	0,0005	0,0225
> 12	0	0,0001	0,0044

Como já foi falado a estatística de teste é χ^2 dada por

$$\chi^2 = \sum_{i=0}^{k} \frac{(o_i - e_i)^2}{e_i}$$

$$= \frac{(1-1,5592)^2}{1,5592} + \frac{(12-5,2078)^2}{5,2078} + ... + \frac{(0-0,0044)^2}{0,0044} = 48,7508,$$

e a variável de interesse é a forma da distribuição do número de reclamações sobre os médicos.

Testa-se as seguintes hipóteses

H_0 : a amostra segue uma distribuição Poisson.

H_1 : a amostra não segue uma distribuição Poisson.

Adota-se $\alpha = 0.05$ e rejeita-se H_0 se $\chi^2_{calc} > \chi^2_{tab}$.

Pela construção teórica o número de graus de liberdade da estatística de teste χ^2 é dada por $k - p - 1 = 11$, onde p é o número de parâmetros estimado.

Com os cálculos já realizados, segue-se que $\chi^2_{calc} = 48,75 > \chi^2_{tab} = 19,675$, daí rejeita-se H_0, ou seja, existem evidências que o conjunto de dados não segue uma distribuição de Poisson(3,34).

3.2 Aplicação 2

Agora, utiliza-se dados simulados para ilustrar a aproximação da Binomial pela Poisson. Utilizando-se o R para gerar os números aleatórios, adota-se a codificação "r" gerador de números aleatórios, seguida da função de distribuição, ou seja, "$rbinom$" . Gera-se três amostras com probabilidades distintas, ou seja,

$> rbinom(30, 10, 0.1)$
1 4 1 1 2 0 2 1 1 0 0 0 2 2 0 1 1 0 3 2 1 1 2 0 0 0 1 0 0 2

Verificando-se, através da distribuição de Poisson com parâmetro $\lambda = n \times p = 10 \times 0.1 = 1$, calcula-se as probabilidades, os valores esperados e em seguida faz-se o teste para verificação do modelo. Observe a Tabela 3.

Tabela 3- os valores observados na amostra de tamanho 30 e parâmetros (10, 0.1).

X	o_i	p_i	$e_i = np_i$
0	11	$0,3678$	$11,034$
1	10	$0,3678$	$11,034$
2	07	$0,1839$	$5,517$
3	01	$0,0613$	$1,839$
4	01	$0,0153$	$0,459$
> 5	00	$0,0039$	$0,117$

A estatística de teste é

$$\chi^2 = \sum_{i=0}^{k} \frac{(o_i - e_i)^2}{e_i}$$

$$= \frac{(11 - 11,034)^2}{11,034} + ... + \frac{(0 - 0,117)^2}{0,117} = 2,682$$

Logo testa-se as hipóteses

H_0: os números simulados de uma distribuição Binomial (n, p) é aproximadamente uma distribuição de Poisson (λ).

H_1: os números simulados de uma distribuição Binomial (n, p) não é aproximadamente uma distribuição de Poisson(λ).

Usa-se $\alpha = 0.05$ e rejeita-se H_0 se $\chi^2_{calc} > \chi^2_{tab}$.

Com os cálculos feito a mão, basta usar a tabela de distribuição qui-quadrado com 4 graus de liberdades, com $\alpha = 0.05$, segue que $\chi^2_{calc} = 2,682 < \chi^2_{tab} = 9,488$, daí aceita-se H_0, ou seja, o conjunto de dados simulados segue uma distribuição de Poisson(1).

Verificando-se a segunda amostra, segui-se o mesmo raciocinio da primeira.

$> rbinom(30, 50, 0.05)$

4 2 1 1 1 3 1 1 0 4 2 2 3 2 2 3 6 5 2 3 4 0 2 2 0 3 4 3 1 3

$$\lambda = n \times p = 50 \times 0,05 = 2,5$$

Tabela 4: Os valores observados na amostra de tamanho 30 e parâmetros (50, 0.05).

X	o_i	p_i	$e_i = np_i$
0	03	$0,082$	$2,46$
1	06	$0,205$	$6,15$
2	08	$0,257$	$7,71$
3	07	$0,214$	$6,42$
4	04	$0,134$	$4,02$
5	01	$0,067$	$2,01$
6	01	$0,028$	$0,84$
> 7	00	$0,011$	$0,33$

A estatística de teste é

$$\chi^2 = \sum_{i=0}^{k} \frac{(o_i - e_i)^2}{e_i}$$

$$= \frac{(3 - 1,2,46)^2}{2,46} + ... + \frac{(0 - 0,33)^2}{0,33} = 1,87$$

$$\chi^2_{calc} = 1,087 < \chi^2_{tab} = 12,592$$

Logo testa-se as hipóteses

H_0: os números simulados de uma distribuição Binomial (n, p) é aproximadamente uma distribuição de Poisson (λ).

H_1: os números simulados de uma distribuição Binomial (n, p) não é aproximadamente uma distribuição de Poisson(λ).

Usa-se $\alpha = 0,05$ e rejeita-se H_0 se $\chi^2_{calc} > \chi^2_{tab}$.
Com os calculos feito a mão, basta usar a tabela de distribuição qui-quadrado com 6 graus de liberdades, com $\alpha = 0,05$, segue que $\chi^2_{calc} = 1,087 < \chi^2_{tab} = 12,592$, daí aceita-se H_0, ou seja, o conjunto de dados simulados segue uma distribuição de Poisson(2,5).

Idem

$> rbinom(30, 50, 0.005)$

0 0 0 2 1 1 0 1 0 1 0 0 0 1 1 0 0 1 0 0 0 1 0 1 1 1 0 0 0 0

$$\lambda = n \times p = 50 \times 0,005 = 0,25$$

Tabela 5: Os valores observados na amostras de tamanho 30 e parâmetros (50, 0.005).

X	o_i	p_i	$e_i = np_i$
0	18	$0,779$	$23,34$
1	11	$0,195$	$5,85$
2	01	$0,024$	$0,72$
> 3	00	$0,002$	$0,06$

A estatística de teste é

$$\chi^2 = \sum_{i=0}^{k} \frac{(o_i - e_i)^2}{e_i}$$

$$= \frac{(18 - 23,34)^2}{23,34} + ... + \frac{(0 - 0,06)^2}{0,06} = 5,92$$

$$\chi^2_{calc} = 5,92 < \chi^2_{tab} = 5,991$$

Logo testa-se as hipóteses

H_0: os números simulados de uma distribuição Binomial (n, p) é aproximadamente uma distribuição de Poisson (λ).

H_1: os números simulados de uma distribuição Binomial(n, p) não é aproximadamente uma distribuição de Poisson(λ).

Usa-se $\alpha = 0.05$ e rejeita-se H_0 se $\chi^2_{calc} > \chi^2_{tab}$.
Com os cálculos feito a mão, basta usar a tabela de distribuição qui-quadrado com 2 graus de liberdades, com $\alpha = 0,05$, segue que $\chi^2_{calc} = 5,92 < \chi^2_{tab} = 5,991$, daí aceita-se H_0, ou seja, o conjunto de dados simulados segue uma distribuição de Poisson(0,25).

4 Conclusão

A distribuição Poisson desempenha um papel importante nas tomadas de decisão, encontrando uma vasta área de aplicação. Frequentemente usa-se distribuições para cálculo de probabilidades associados aos resultados de interesse. Com o avanço da tecnologia e das ferramentas computacionais se tornou mais fácil fazer simulações de amostras para inferir em relação a tal distribuição. Assim pode-se relacionar a teoria com a prática, dando significado a aprendizagem.

Este trabalho teve por objetivo fazer uma revisão teórica da distribuição de Poisson. Com duas aplicações, uma através do conjunto de dados disponível no R e outra através da simulação dos números aleatórios também feitos no R. Através do comando "$rbinom(x, n, p)$" gera-se três amostras da distribuição binomial com probabilidades distintas e em seguida testamos o modelo usando o teste qui-quadrado. No conjunto de dados observados no R ver-se que, mesmo sendo um processo de contagem não mostrou seguir uma distribuição Poisson. Nas amostras simuladas da distribuição binomial, mostrou-se que a variável aleatória de poisson pode ser usada como aproximação para a variável aleatória binomial.

Referências Bibliograficas

BUSSAB,W.O.; MORETTIN,P.A. *Estatística Básica.* 5. ed. São Paulo: Saraiva, 2005.

MEYER,P. L. *Probabilidade: Aplicações à Estatística.*2.ed. Rio de Janeiro: LTC,1983.

MOOD,A.M.; GRAYBILL,A.F. *Introduction to the Theory of Statistics.* Library of Congress Cataloging in Publication Data. New York, 1963

PETERNELLI,L.A.;MELO,M.P *Conhecendo o R: Uma visão Estatística.* Ed. UFV. Viçosa, MG, 2011.

ROSS,S. *PROBABILIDADE: Um curso moderno com aplicações*/ Sheldon Ross; Tradutor: Albert Resend de Conti.-8.ed.-Porto Alegre: Bookman,2010.608p.