Amostras Aleatórias e Distribuições Amostrais

Probabilidade e Estatística: afinal, qual é a diferença?

Até agora o que fizemos foi desenvolver ***modelos probabilísticos*** que se adequavam a situações reais. Por exemplo, indicamos quando os modelos Binomial, Poisson, Exponencial, Normal, Uniforme, etc ... eram adequados. Todos estes modelos referem-se a distribuições de probabilidade que envolvem ***parâmetros***, que ***até agora*** foram ***supostos conhecidos***. Para que as probabilidades associadas a eventos sejam calculadas é necessário conhecer o valor destes parâmetros.

No estudo das probabilidades, o nosso objetivo é calcular a probabilidade de eventos pré-especificados. ***De agora em diante*** teremos um novo objetivo. A partir de uma ***amostra*** de uma distribuição de probabilidade especificada pretendemos aprender alguma coisa sobre os parâmetros da distribuição, isto é, estaremos interessados em ***estimar os parâmetros*** da distribuição de probabilidade.

Esta é a grande diferença entre Probabilidade e Estatística. No estudo de Probabilidade estamos interessados em definir modelos que possam ser aplicados a situações reais. Estes modelos envolvem distribuições de probabilidade totalmente conhecidas, isto é, não apenas a forma da densidade, mas também os seus parâmetros são conhecidos. No estudo da Estatística supõe-se que o modelo probabilístico é conhecido, isto é, sabe-se qual a distribuição de probabilidade que modela a situação real, mas os parâmetros desta distribuição são desconhecidos, e devem ser estimados a partir dos dados.

O nosso objetivo em Estatística é descobrir alguma coisa sobre os parâmetros desconhecidos de uma distribuição de probabilidade. Os mecanismos mais usuais para "inferir" alguma coisa sobre estes parâmetros são:

1) *Estimação pontual* - o objetivo é "chutar" os valores do parâmetro desconhecido.

2) *Estimação por intervalos* - o objetivo é encontrar um intervalo que contenha o parâmetro de interesse com uma probabilidade especificada.

3) *Testes de hipóteses* - o objetivo é criar conjecturas sobre os valores possíveis do parâmetro e verificar se estas conjecturas são muito ou pouco prováveis (isto é, testar as hipóteses).

Todos estes procedimentos são baseados na noção de **amostra aleatória**.

Definição (amostra, ou amostra aleatória)
Uma amostra aleatória é um conjunto de variáveis aleatórias **independentes e identicamente distribuídas** (iid).

Notação : a.a. = amostra aleatória

O que se faz na prática?
Para ganhar informação sobre os parâmetros desconhecidos de uma distribuição de probabilidade usamos um conjunto de variáveis aleatórias independentes e identicamente distribuídas. Isto equivale a repetir a experiência aleatória que está sendo descrita pelo modelo em questão n vezes, em condições idênticas e de maneira independente. A partir dos valores observados das variáveis X_1, X_2, ..., X_n calcularemos funções que nos permitirão aprender sobre os parâmetros desconhecidos do modelo. Estas funções serão chamadas de "estatísticas".

Definição (estatística)
Seja X_1, X_2, ..., X_n uma a.a. de uma variável aleatória X. Sejam x_1, x_2, ..., x_n os valores observados de X_1, X_2, ..., X_n .

Seja $Y = h(X_1, X_2, ..., X_n)$ uma função apenas das variáveis X_1, X_2, ..., X_n . Y é chamado de "**estatística**".

Note que *uma estatística não é função de parâmetros desconhecidos*, ela só envolve as variáveis na amostra aleatória, ou seja, pode ser diretamente computada a partir dos valores observados numa amostra.

Por definição, qualquer estatística Y é uma variável aleatória, e tem uma distribuição de probabilidade que depende da distribuição de X_1, X_2, ..., X_n .

O nosso problema então é encontrar estatísticas que sirvam como bons estimadores pontuais de parâmetros desconhecidos. Também é importante definir critérios que nos permitam dizer que uma estatística é "melhor" que outra para estimar um dado parâmetro.

De uma maneira geral, as estatísticas devem conter "toda" a informação presente numa amostra. Se não fosse assim, não valeria a pena calcular uma estatística, a gente simplesmente usaria uma única observação da amostra. Este acréscimo de informação representado pelo uso de uma estatística (*ao invés de uma única* observação) geralmente se traduz por uma considerável redução na variância. Por exemplo, a variância da média amostral é igual à variância de cada observação dividida pelo tamanho da amostra. Quanto maior o tamanho da amostra, menor é a variância da média amostral, isto é, mais "precisa" é a média amostral.

As estatísticas mais famosas

Sejam X_1, X_2, ..., X_n uma amostra aleatória de uma distribuição qualquer. As estatísticas mais comuns, calculadas a partir desta amostra são:

1) Média amostral

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

2) Variância amostral

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$$

3) Desvio padrão amostral

$$S = \sqrt{S^2} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}$$

4) Mínimo da amostra

$$X_{(1)} = min\left(X_1, X_2, \ldots, X_n\right)$$

5) Máximo da amostra

$$X_{(n)} = max\left(X_1, X_2, \ldots, X_n\right)$$

6) Amplitude da amostra

$$A = X_{(n)} - X_{(1)}$$

7) k-ésima estatística de ordem

É o k-ésimo elemento da amostra ordenada. Por exemplo, $X_{(2)}$ é o segundo menor elemento da amostra $X_1, X_2, ..., X_n$.

Um dos nossos objetivos aqui é desenvolver as distribuições de estatísticas obtidas a partir de uma amostra aleatória da distribuição Normal.

O próximo teorema refere-se à média amostral de uma amostra aleatória da densidade Normal.

Teorema

Sejam $X_1, X_2, ..., X_n$ uma amostra aleatória da distribuição $N(\mu, \sigma^2)$. Seja $\overline{X}$ a média amostral. Então:

$$\overline{X} \approx N\left(\mu, \frac{\sigma^2}{n}\right)$$

A demonstração do teorema é trivial , e segue das propriedades da função geradora de momentos.

Este teorema pode ser generalizado para uma amostra aleatória de uma distribuição qualquer.

Teorema

Sejam $X_1, X_2, ..., X_n$ uma amostra aleatória de uma distribuição qualquer tal que $E(X_i) = \mu$ e $VAR(X_i) = \sigma^2$. Seja $\overline{X}$ a média amostral. Então:

1) $E(\overline{X}) = \mu$

2) $VAR(\overline{X}) = \sigma^2 / n$

3) Se n é grande, pelo teorema central do limite podemos concluir que:

$$\sqrt{n}.\frac{\left(\overline{X} - \mu\right)}{\sigma}$$

é aproximadamente N(0,1).

Note que, neste caso, nada é dito a respeito da distribuição de X. Apenas a sua média e variância são conhecidas, e são funções da média e variância de cada X_i.

A princípio a distribuição de $\overline{X}$ poderia ser uma coisa estranha, que não tem nada a ver com a distribuição original de cada X_i. No entanto, se o tamanho da amostra é grande podemos concluir que a distribuição de $\overline{X}$, devidamente escalonada, é aproximadamente N(0,1).

O próximo teorema refere-se à distribuição do máximo e do mínimo de uma amostra.

Teorema

Sejam $X_1, X_2, ..., X_n$ uma amostra aleatória de uma distribuição ***contínua*** qualquer com densidade f(.) e função de distribuição F(.). Sejam $X_{(1)}$ e $X_{(n)}$ respectivamente, o mínimo e o máximo da amostra. Então as densidades de $X_{(1)}$ e $X_{(n)}$ são dadas por:

1) Densidade do mínimo

$$g_1(x) = n.f(x).(1 - F(x))^{n-1}$$

2) Densidade do máximo

$$g_n(x) = n.f(x).(F(x))^{n-1}$$

Demonstração

Só faremos a demonstração do segundo item (máximo da amostra). A demonstração do outro item é semelhante.

Note que se $X_{(n)}$ é o máximo da amostra, então $X_{(n)} < k$ equivale a : todo $X_i < k$, para qualquer número k. Logo, a função de distribuição do máximo pode ser facilmente encontrada, e é dada por:

$$G_n(k) = \Pr\left(X_{(n)} \le k\right) = \Pr\left(X_1 \le k, X_2 \le k, \ldots., X_n \le k\right)$$

Também, os X_i 's são independentes, e esta última probabilidade pode ser escrita como o produto das probabilidades para cada X_i . Então:

$$G_n(k) = \Pr\left(X_1 \le k, X_2 \le k, \ldots., X_n \le k\right) = \Pr\left(X_1 \le k\right).\Pr\left(X_2 \le k\right)\ldots\Pr\left(X_n \le k\right)$$

Como os X_i são identicamente distribuídos, estas probabilidades são as mesmas para todo X_i e correspondem à função de distribuição F(.) com argumento k.

$$G_n(k) = \left(\Pr\left(X_1 \le k\right)\right)^n = \left(F(k)\right)^n$$

A densidade de $X_{(n)}$ é encontrada derivando-se a função de distribuição com relação ao argumento k, e lembrando que a derivada de F(.) é f(.), a densidade de cada X_i . Então :

$$g_n(k) = \frac{dG_n(k)}{dk} = n.\left(F(k)\right)^{n-1}.\frac{dF(k)}{dk} = n.f(k).\left(F(k)\right)^{n-1}$$

Exemplo

Sejam $X_1, X_2, ..., X_n$ uma amostra aleatória da densidade Exponencial com parâmetro λ. Encontre a densidade de $X_{(1)}$, o mínimo da amostra.

Solução

A densidade de cada X_i é:

$$f(x) = \lambda.e^{-\lambda x}$$

A função de distribuição é:

$$F(x) = \Pr\left(X \le x\right) = \int_0^x \lambda.e^{-\lambda t}dt = 1 - e^{-\lambda x}$$

A densidade do mínimo é, pelo teorema anterior:

$$g_1(y) = n.\left(1 - F(y)\right)^{n-1}.f(y) = n.\left(1 - 1 + e^{-\lambda y}\right)^{n-1}.\left(\lambda.e^{-\lambda y}\right) =$$

$$= n.\lambda.\left(e^{-\lambda y}\right)^{n-1+1} = n.\lambda.e^{-n.\lambda.y}$$

Ou seja, $X_{(1)}$ tem densidade Exponencial com parâmetro $n.\lambda$.

Exemplo

A duração de um componente eletrônico é uma variável aleatória T com distribuição Exponencial com parâmetro $\lambda = 0.001$.

Testou-se 100 componentes e observou-se a duração de cada um deles, gerando uma amostra aleatória $T_1, T_2, ..., T_{100}$.

Calcule as seguintes probabilidades:

a) Pr (950 < T < 1100)

b) Pr (W > 7200) onde W = máx(T_1, T_2 , ..., T_{100})

c) Pr (V < 10) onde V = mín(T_1, T_2 , ..., T_{100})

Solução

a) Note que, se T_i ~ Expo(0.001) para i =1 , 2, ..., 100 então :

$E(T_i) = 1/ 0.001 = 1000$ e

$VAR(T_i) = 1/(0.001)^2 = 10^6$

Assim:

$E(\overline{T}) = E(T_i) = 1000$ e $VAR(\overline{T}) = VAR(T_i)/100 = 10^4$

Pelo teorema central do limite:

$$Z = \frac{\overline{T} - 1000}{\sqrt{10^4}} = \frac{\overline{T} - 1000}{100}$$

tem aproximadamente a distribuição N(0,1) .

Assim:

$$\Pr\left(950 \le \overline{T} \le 1100\right) = \Pr\left(\frac{950-1000}{100} \le \frac{\overline{T}-1000}{100} \le \frac{1100-1000}{100}\right) =$$

$$= \Pr(-0.5 \le Z \le 1) = \Phi(1) - \Phi(-0.5) = 0.532$$

Onde estas últimas probabilidades foram obtidas da tabela N(0,1).

b) Pr (W > 7200) = Pr{ máx(T_1, T_2 , ..., T_{100}) > 7200 } =

= 1 - Pr{ máx(T_1, T_2 , ..., T_{100}) ≤ 7200 }

Mas, se W = máx(T_1, T_2 , ..., T_{100}) ≤ 7200 então todos os T_i são ≤ 7200.

$$\Pr(W \le 7200) = \Pr\left(T_1 \le 7200, T_2 \le 7200, \ldots, T_{100} \le 7200\right) =$$

$$= \left(\Pr\left(T_1 \le 7200\right)\right)^{100} = \left(1 - e^{-0.001(7200)}\right)^{100} = \left(1 - e^{-7.2}\right)^{100} = (0.99925)^{100} = 0.928$$

c) Pr (V < 10) onde V = mín(T_1, T_2 , ..., T_{100})

Pr (V < 10) = 1 - Pr(V ≥ 10) = 1- Pr(mín(T_1, T_2 , ..., T_{100}) ≥ 10)

Mas, se mín(T_1, T_2 , ..., T_{100}) ≥ 10 então todos os T_i também são ≥ 10.

Logo, Pr(V < 10) = 1 - Pr(T_1 ≥ 10, T_2 ≥ 10, ..., T_{100} ≥ 10) =

$$= 1 - \left[\Pr\left(T_1 \ge 10\right)\right]^{100} = 1 - \left[e^{-0.001(10)}\right]^{100} = 1 - \left[e^{-0.01}\right]^{100} = 1 - e^{-1} = 0.632$$

A distribuição Qui-Quadrado

Definição (densidade Qui-Quadrado com k graus de liberdade)

Seja X uma variável aleatória contínua e positiva com densidade dada por:

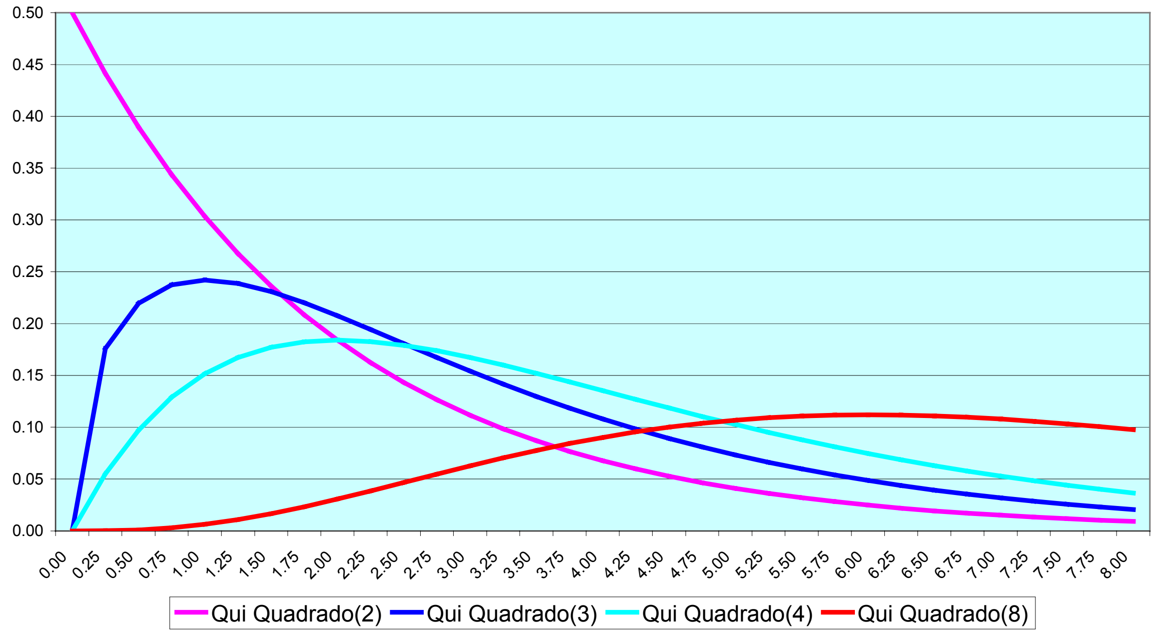
$$f(x) = \frac{1}{2^{k/2}.\Gamma\left(\frac{k}{2}\right)}.x^{\frac{k}{2}-1}.e^{-x/2} \quad \text{onde } x > 0$$

Então X tem densidade Qui-Quadrado com k graus de liberdade, e escrevemos : $X \sim \chi^2_k$

A densidade Qui-Quadrado com k graus de liberdade é apenas um caso particular da densidade Gama. Na verdade:

$$\chi_k^2 \quad = \quad \text{Gama}(\ \alpha = k/2,\ \beta = 1/2)$$

Densidades Qui-Quadrado com 2, 3, 4 e 8 Graus de Liberdade



Teorema

Se X tem densidade Qui-Quadrado com k graus de liberdade então sua média, variância e função geradora de momentos são dadas por:

M. Barros Consultoria Ltda.
e-mail: barrosm@alumni.utexas.net
info@mbarros.com

E(X) = k

VAR(X) = 2.k

$$M(t) = \frac{1}{(1-2t)^{r/2}}$$

Demonstração

Segue direto dos resultados correspondentes para a densidade Gama.

A densidade Qui-Quadrado é tabelada. As tabelas desta densidade fornecem os pontos tais que a probabilidade da variável estar acima deles é especificada. Uma pequena porção de uma tabela da densidade Qui-Quadrado é mostrada a seguir.

graus de liberdade ↓	0.990	0.950	0.050	0.01
2	0.020	0.100	5.99	9.21
6	0.870	1.640	12.59	16.81
12	3.570	5.23	21.03	26.22

Por exemplo:

Supondo que X seja uma variável aleatória com densidade Qui-Quadrado com 6 graus de liberdade, a probabilidade de X exceder 0.87 é 99%. Analogamente, a probabilidade de X exceder 12.59 é 5% e a probabilidade de X estar acima de 16.81 é apenas 1%.

Uma propriedade muito importante da densidade Qui-Quadrado é a preservação da mesma família de densidades quando somamos variáveis independentes. Ou seja, se X_1, X_2, ..., X_n são variáveis **independentes**, cada uma com distribuição Qui-Quadrado, a soma de X_1, X_2, ..., X_n também é uma variável aleatória Qui-Quadrado.

Teorema (aditividade da densidade Qui-Quadrado)

Sejam X_1, X_2, ..., X_n variáveis aleatórias independentes, e suponha que X_i tem densidade Qui-Quadrado com k_i graus de liberdade. Seja $Y = X_1 + X_2 + ... + X_n$. Então Y tem também uma densidade Qui-Quadrado, mas com $k = k_1 + k_2 + ... + k_n$ graus de liberdade.

O próximo teorema exibe a relação existente entre as densidades Normal padrão e Qui-Quadrado.

Teorema

Seja Z ~ N(0,1) . Então V = Z^2 tem densidade Qui-Quadrado com 1 grau de liberdade.

Demonstração

A demonstração é feita usando-se o método da função de distribuição, já que a função V = Z^2 não é injetora, o que nos impede de usar o método do jacobiano :

G(v) = Pr(V ≤ v) = Pr(Z^2 ≤ v) = Pr(- √v ≤ Z ≤ +√v) = Φ(+√v) - Φ(-√v)

onde Φ(.) indica a função de distribuição de uma variável aleatória N(0,1).

Derivando esta expressão em relação a v resulta na densidade de V, que é :

$$g(v) = \left(\frac{1}{\sqrt{2\pi}} . \exp\left(-\frac{\left(\sqrt{v}\right)^2}{2} \right) \right) . \left(\frac{1}{2} . v^{-1/2} \right) - \left(\frac{1}{\sqrt{2\pi}} . \exp\left(-\frac{\left(-\sqrt{v}\right)^2}{2} \right) \right) . \left(\frac{-1}{2} . v^{-1/2} \right) =$$

$$= \left(\frac{1}{2} . v^{-1/2} \right) . \left(\frac{2}{\sqrt{2\pi}} \exp\left(-\frac{v}{2} \right) \right) = \frac{1}{\sqrt{2\pi}} v^{-1/2} . e^{-v/2}$$

Isto é :

$$g(v) = \frac{1}{2^{1/2}\sqrt{\pi}} v^{\frac{1}{2}-1} . e^{-v/2} = \frac{1}{2^{1/2}\Gamma\left(\frac{1}{2}\right)} v^{\frac{1}{2}-1} . e^{-v/2}$$

Substituindo k = 1 na definição da densidade Qui-Quadrado resulta na expressão acima, o que prova o teorema.

A combinação dos 2 últimos teoremas leva a um resultado importante.

Teorema

Sejam Z_1, Z_2, ..., Z_n variáveis aleatórias independentes e identicamente distribuídas com densidade N(0,1). Então:

$$V = \sum_{i=1}^{n} Z_i^2 = Z_1^2 + Z_2^2 + \ldots + Z_n^2$$

tem densidade Qui-Quadrado com n graus de liberdade.

Este resultado segue trivialmente dos dois últimos teoremas, se lembrarmos que cada Z_i^2 tem densidade Qui-Quadrado com 1 grau de liberdade (e são todos independentes).

Por que a densidade Qui-Quadrado é importante?
Esta densidade está relacionada com a distribuição da variância amostral obtida a partir de uma amostra aleatória Normal, como indicado no próximo teorema.

Teorema

Sejam $X_1, X_2, ..., X_n$ uma amostra aleatória da distribuição $N(\mu, \sigma^2)$. Seja S^2 a variância amostral, dada por:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$$

Então:

$$\frac{(n-1)S^2}{\sigma^2} = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}{\sigma^2}$$

tem distribuição Qui-Quadrado com (n-1) graus de liberdade.

A partir deste teorema podemos deduzir facilmente a média e variância de S^2.

Teorema

Sejam $X_1, X_2, ..., X_n$ uma amostra aleatória da distribuição $N(\mu, \sigma^2)$. Seja S^2 a variância amostral. Então :

$$E(S^2) = \sigma^2$$

$$VAR(S^2) = \frac{2\sigma^4}{n-1}$$

Demonstração

Pelo teorema anterior e sabendo a média e variância de uma variável aleatória Qui-Quadrado temos:

$$E\left(\frac{(n-1)S^2}{\sigma^2}\right) = n-1 \Rightarrow E\left(S^2\right) = \frac{(n-1)\sigma^2}{(n-1)} = \sigma^2$$

$$VAR\left(\frac{(n-1)S^2}{\sigma^2}\right) = 2.(n-1) \Rightarrow VAR\left(S^2\right) = \frac{2.(n-1).\left(\sigma^2\right)^2}{(n-1)^2} = \frac{2.\sigma^4}{n-1}$$

M. Barros Consultoria Ltda.
e-mail: barrosm@alumni.utexas.net
info@mbarros.com