Inferência Estatística

Aula 2

Agosto de 2008

Mônica Barros

Conteúdo – Revisão de Probabilidade

- Algumas das principais distribuições discretas
  - Distribuição de Bernoulli
  - Distribuição Binomial
  - Distribuição Hipergeométrica
  - Distribuição Geométrica

Nota

- A seguir apresentaremos diversas distribuições discretas.
- Serão mostradas também suas médias, variâncias e funções geradoras de momentos, embora estes tópicos só sejam abordados na próxima aula. Você deverá anotar estas quantidades, pois elas serão úteis no restante do curso.

Distribuição Bernoulli

- É a mais simples v.a. discreta.
- Seja X uma v.a. com apenas dois valores possíveis, “sucesso” (denotado por 1) e “falha” (denotado por 0).
- Então:

$$f(1) = \Pr(X = 1) = p$$
$$f(0) = \Pr(X = 0) = 1 - p = q$$

- Note que 0 < p < 1 e “sucesso” e “falha” não indicam se o resultado de uma experiência é “bom”ou “ruim”.

Distribuição Bernoulli

- A distribuição de Bernoulli serve como um "tijolo" para a construção de modelos mais elaborados, como a Binomial, a Geométrica e a Binomial Negativa.
- Podemos reescrever a função de probabilidade como:

$$\boxed{f(x) = \Pr(X = x) = p^x (1-p)^{1-x} \quad \text{onde } x = 0,1}$$

- Esta última notação será útil para identificar uma variável Bernoulli apenas como um caso particular de uma Binomial.
- Notação: X ~ Bernoulli(p)

Distribuição Binomial

- É importante no contexto de amostragem COM reposição, mas seu uso é muito mais geral.
- A situação clássica em que usamos uma variável Binomial é:
  - Uma experiência aleatória tem apenas dois resultados possíveis: "sucesso" e "falha", onde a probabilidade de "sucesso" é <u>p</u> e a probabilidade de "falha" é q = 1- p.

Distribuição Binomial

- A experiência é repetida um número fixo (n) de vezes, sempre nas mesmas condições, de tal forma que as probabilidades de "sucesso" (p) e "falha" (q =1-p) se mantêm inalteradas a cada repetição.
- As diversas repetições da experiência são feitas de maneira independente, ou seja, o resultado de uma repetição não afeta o resultado das outras.

Distribuição Binomial

- A variável aleatória X que mede o número de "sucessos" nas n repetições da experiência é uma variável discreta, com valores possíveis 0, 1, 2, ... , n.
- Dizemos que esta variável tem distribuição Binomial com parâmetros n e p, e escrevemos X ~ Bin(n, p).

Distribuição Binomial - quadro resumo

experiência aleatória	"sucesso"	"falha"	p = probabilidade de "sucesso"	n = número de repetições da experiência	X = variável aleatória Binomial
"chutar" a resposta numa prova de múltipla escolha onde cada questão tem 5 opções	acertar a resposta da questão	errar a resposta	1/5	número de questões da prova	número de respostas certas na prova
nascimento de uma criança numa família	menina	menino	1/2	número de crianças na família	número de meninas na família
jogada de um dado	sair o número 6	sair qualquer outro número	1/6	número de jogadas do dado	número de vezes em que saiu o número 6 nas n jogadas do dado
verificar se uma peça produzida numa fábrica tem defeito	peça tem defeito	peça não tem defeito	proporção de peças com defeito na população de peças	tamanho da amostra	número de peças com defeito na amostra

Distribuição Binomial

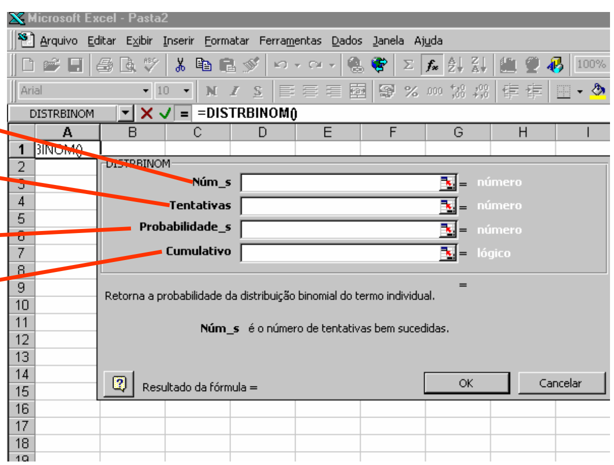
- Se X ~ Bin(n,p), sua função de probabilidade é:

$$f(x) = \Pr(X = x) = \binom{n}{x} p^x (1-p)^{n-x} = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} \quad \text{para } x = 0,1,2,\ldots,n$$

- No Excel:
  - Use a função estatística *distrbinom* (ou *dist.binom*, dependendo da versão do Excel).

Distribuição Binomial

- No Excel:



Valor de X

Valor de n

Valor de p

Argumento lógico - se VERDADEIRO produz a função de distribuição acumulada, se FALSO produz a função de probabilidade

Distribuição Binomial

- Exemplo: em uma loja, a probabilidade de um cliente realizar uma compra é de 15%. Qual a probabilidade de, entre 5 clientes que entram na loja, exatamente 3 realizarem uma compra?

$$n = 5; \quad x = 3; \quad p = 0.15$$

$$f(3) = \frac{5!}{3!.2!}(0.15)^3.(0.85)^2 = 0.024$$

Distribuição Binomial

- **Exemplo**

Numa eleição supõe-se que 30% dos eleitores são favoráveis a uma certa proposta. Toma-se uma amostra de tamanho 20 de eleitores da cidade do Rio de Janeiro. Calcular as probabilidades de 4, 5, 6, 7, 8, 9 ou 10 dos eleitores na amostra serem favoráveis à proposta.

- **Solução**

Seja X o número de eleitores na amostra que são a favor da proposta. Então os valores possíveis de X são 0, 1, 2, ..., 20, e X tem distribuição Binomial com parâmetros n = 20 e p = 0.30.

Distribuição Binomial

- **As probabilidades para os diversos valores de X são calculadas através da fórmula:**

$$\Pr(X = x) = \binom{20}{x}.(0.30)^x.(0.70)^{20-x} = \frac{20!}{x!(20-x)!}.(0.30)^x.(0.70)^{20-x}$$

- **A tabela a seguir foi produzida usando a função distrbinom do Excel para x = 4, 5, ..., 10.**

x	Pr(X =x)
4	13.04%
5	17.89%
6	19.16%
7	16.43%
8	11.44%
9	6.50%
10	3.08%

Exemplo (para casa – use o Excel)

- **Uma empresa aérea sabe que 20% das pessoas que fazem reservas aéreas cancelam suas reservas.**
- **A empresa vende 50 passagens para um vôo que contém apenas 46 lugares. Supondo que as pessoas cancelam suas reservas de maneira independente, calcule a probabilidade de que haverá assentos para todos os passageiros.**

Exemplo (para casa – use o Excel)

- **Você arranjou um emprego numa pizzaria que funciona no sistema de entrega a domicílio. Apenas 5% dos pedidos são de pizza de lombinho com abacaxi.**

Você recebe exatamente 9 pedidos pelo telefone, qual a probabilidade de, no máximo, 1 pizza de lombinho com abacaxi ser pedida?

Você recebe exatamente 30 pedidos pelo telefone num dia de bastante movimento, qual a probabilidade de receber mais de 3 pedidos de pizza de lombinho com abacaxi? Dica: pare e pense antes de fazer contas desnecessárias!!!!!!

Distribuição Hipergeométrica

- A distribuição hipergeométrica surge no contexto de amostragem SEM reposição em que o tamanho da população é muito maior que o da amostra, de tal forma que a densidade Binomial não pode ser usada.
- Sejam N e n respectivamente os tamanhos da população e da amostra, e suponha que na população existem r objetos do tipo A e N-r objetos do tipo B.
- Seja X a v.a. que indica o número de objetos do tipo A na amostra. Então a probabilidades de exatamente x objetos do tipo A na amostra é:

Distribuição Hipergeométrica

$$f(x) = \Pr(X = x) = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}$$

onde x = 0, 1, 2, ..., mín(n, r).

- Esta função de probabilidade também está implementada no Excel, como veremos na próxima figura.

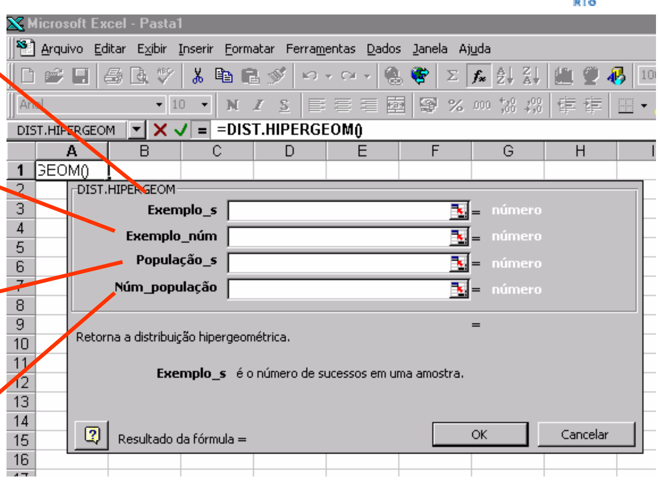
Distribuição Hipergeométrica

Valor de X

Tamanho da amostra (n)

Número de objetos do tipo A na população (r)

Tamanho da população (N)



Distribuição Hipergeométrica

- **Exemplo**
- Uma empresa encomenda 50 placas fax modem de um fornecedor novo para montagem nos seus micros. A probabilidade de defeito numa placa é 10% . A empresa toma uma amostra de 10 micros sem reposição. Qual a probabilidade de 4 deles apresentarem uma placa fax modem com defeito?
- **Solução**
- Seja X o número de micros com placa defeituosa na amostra. Os valores possíveis de X são 0, 1, 2, 3, 4, 5 = mín(5,10) Procuramos calcular Pr(X=4).

Distribuição Hipergeométrica

$$\Pr(X = 4) = f(4) = \frac{\binom{5}{4}\cdot\binom{45}{6}}{\binom{50}{10}} = \frac{5.(45!).(10!).(40!)}{(6!)(39!)(50!)} =$$

$$= \frac{200(45!)(10!)}{(6!)(50!)} = \frac{200(45!)(10)(9)(8)(7)(6!)}{(6!)(50)(49)(48)(47)(46)(45!)} =$$

$$= \frac{200(10)(9)(8)(7)}{(50)(49)(48)(47)(46)} = 0.00396 = 0.396\%$$

Distribuição Hipergeométrica

Exemplo (P1 – semestre 2001.02)

- **Uma empresa de cartão de crédito considera aceitável que 3% das chamadas recebidas pelo seu serviço de atendimento ao consumidor sejam reclamações sobre faturamento.**

- **Para verificar se os serviços prestados estão "sob controle" faz-se uma pesquisa a partir de 40 ligações recebidas. Suponha que neste grupo existem, na verdade, 4 reclamações sobre faturamento.**

Distribuição Hipergeométrica

- **A empresa utiliza o seguinte processo para verificar a qualidade dos seus serviços: *a partir de uma amostra de 5 ligações,* se *no máximo 1 ligação refere-se a reclamações sobre faturamento,* a empresa admite que o processo está sob controle, e não toma providências adicionais.**
- **Do contrário, *se mais de uma ligação refere-se a problemas com faturas*, a empresa emite um aviso para a sua gerência financeira, que inicia um processo de verificação das faturas emitidas. Qual a probabilidade da empresa ter de fazer o processo de verificação de faturas, supondo que:**

Distribuição Hipergeométrica

a) A amostragem é feita com reposição (Binomial).

b) A amostragem é feita sem reposição (Hipergeométrica).

Solução

Seja X o número de chamadas com reclamação sobre faturamento na amostra. Então, o processo de verificação de faturas ocorrerá se X > 1, ou seja, se X ≥ 2.

Suponha que o tamanho da população aqui é N = 40 (número de chamadas na pesquisa), e existem r = 4 reclamações sobre faturamento na população.

Distribuição Hipergeométrica

- Então, a proporção de chamadas relativas a faturamento na população é 4/40 = 10% (e portanto bem acima do que seria desejável pelos padrões da empresa, mas ela não sabe disso, a princípio!).
- a) Se a amostragem é feita COM reposição, a distrinuição a ser empregada é Binomial com parâmetros n = 5 e p = 10%. Logo, a probabilidade desejada é:

$$\Pr(X > 1) = \Pr(X \geq 2) = \sum_{x=2}^{5} \frac{5!}{x!(5-x)!}(0.1)^x (0.9)^{5-x}$$

$$= 1 - \sum_{x=0}^{1} \frac{5!}{x!(5-x)!}(0.1)^x (0.9)^{5-x} = 1 - (0.9)^5 - 5(0.1)(0.9)^4 = 0.0815$$

Distribuição Hipergeométrica

- b) Se a amostragem é feita sem reposição, devemos usar a distribuição Hipergeométrica com N = 40, r = 4, N – r = 36, n = 5 e calcular Pr(X > 1) = Pr(X ≥ 2).
- Pelos mesmos argumentos que na situação anterior, é mais conveniente escrever a probabilidade desejada em termos do seu complemento, isto é:
- Pr (X ≥ 2) = 1 – Pr(X=0) – Pr(X=1)

Distribuição Hipergeométrica

- Mas, para qualquer x entre 0 e min(r, 5) = 4

$$\Pr(X = x)\frac{\binom{4}{x}\binom{36}{5-x}}{\binom{40}{5}}$$

- Logo: Pr(X = 0) = 0.5729 e Pr(X = 1) = 0.3581 e portanto Pr (X ≥ 2) = 0.06899.
- Se a amostragem for feita sem reposição, há uma probabilidade de cerca de 6.9% da empresa requerer a verificação do processo de faturamento. Esta probabilidade é aproximadamente 8.2% se a amostragem for com reposição.

Exemplo (para casa)

- Uma empresa de telefonia fixa está preocupada com a qualidade do atendimento prestado aos seus consumidores.
- A empresa considera aceitável que um técnico responsável por um reparo chegue ao local do defeito em 12 horas ou menos (após a reclamação).
- Dados anteriores indicam que há uma probabilidade de 0.7 de um reparo ser efetuado em até 12 horas após a reclamação.

Exemplo (para casa)

- Toma-se uma amostra de 10 consumidores que fizeram uma reclamação sobre defeito na linha telefônica. Calcule as seguintes probabilidades:
- a) De que todos os consumidores serão atendidos dentro do prazo?
- b) De que pelo menos 5 reclamações serão atendidas dentro do prazo?
- c) De que no máximo 3 reclamações serão atendidas FORA do prazo?
- d) Qual a distribuição de probabilidade indicada para modelar este problema e por que?

Distribuição Geométrica

- A distribuição Geométrica, da mesma forma que a Binomial, está relacionada com uma seqüência de tentativas de Bernoulli independentes.
- Ao contrário da distribuição Binomial, a experiência que gera uma variável Geométrica não tem um número de repetições fixo.
- Considere uma seqüência de tentativas de Bernoulli independentes, e seja X uma variável aleatória discreta que mede o número de tentativas necessárias até encontrar o primeiro sucesso.

Distribuição Geométrica

- Um exemplo típico é: imagine uma caixa com um número muito grande de peças. Estas peças podem estar OK ou defeituoscopas, e a probabilidade de encontrar uma peça defeituosa é FIXA e igual a "p".
- Retiram-se peças da caixa até encontrar uma peça com defeito.

Distribuição Geométrica

- Neste exemplo, um "sucesso" indica encontrar uma peça defeituosa.
- Note que estamos supondo que a caixa onde estão as peças contém um número muito grande de objetos, e assim a probabilidade de retirada de uma peça defeituosa é mantida constante (p).

Distribuição Geométrica

- **Seja X o número de repetições necessárias até encontrar uma peça com defeito.**

- **Então a função de probabilidade de X é:**

$$f(x) = \Pr(X = x) = q^{x-1}.p = (1-p)^{x-1}p \quad \text{onde } x = 1,2,3,....$$

- **Já provamos antes que f(x) assim definido é uma função de probabilidade (usando a série geométrica).**

Distribuição Geométrica

- **A função de probabilidade de X é:**

$$f(x) = \Pr(X = x) = q^{x-1}.p \quad \text{onde } x = 1, 2, 3, ...$$

- **Notação: X ~ Geom(p)**
- **Média, Variância e fgm**
  - **Se X ~ Geom(p) então:**

1) E(X) = 1/p

2) VAR(X) = q/p²

$$3) M(t) = \frac{pe^t}{1-qe^t}$$

Distribuição Geométrica

- **Nota: Série Geométrica**

$$\sum_{x=0}^{\infty} a^x = 1 + a + a^2 + ... + a^n + ... = \frac{1}{1-a} \text{ sempre que } |a| < 1$$

- **E se começarmos o somatório em 1:**

$$\sum_{x=1}^{\infty} a^x = a + a^2 + ... + a^n + ... = \frac{1}{1-a} - 1 = \frac{a}{1-a}$$

- **Estas expressões devem ser usadas para demonstrar a função geradora de momentos de uma v.a. Geométrica.**

Distribuição Geométrica - Exemplo

- **Você arranjou um emprego numa empresa que faz pesquisas de opinião pelo telefone.**
- **Apenas 10% das chamadas resultam numa pesquisa completa, isto é, apenas 10% dos entrevistados responde todo o seu questionário. Calcule as seguintes probabilidades:**
- **a) De que a primeira pesquisa completa será respondida na 5a. ligação telefônica.**
- **b) De que a primeira pesquisa completa será respondida na 8a. ligação telefônica.**

Distribuição Geométrica - Exemplo

- Esta é uma aplicação típica da distribuição Geométrica.
- Seja X o número de ligações efetuadas até que a 1a. pesquisa completa seja respondida.
- Então X é uma v.a. Geométrica com probabilidade de sucesso p = 0.10.
- a) $\Pr(X = 5) = (0.90)^4 (0.10) = 6.56\%$
- b) $\Pr(X = 8) = (0.90)^7 (0.10) = 4.78\%$

Distribuição Geométrica

- Exemplo (para casa)
- Um gestor de fundos de investimento ultrapassa a sua meta de retorno mensal 85% das vezes e nos 15% restantes tem um resultado ruim (abaixo da meta).
- Qual a probabilidade dele ter o primeiro resultado ruim nos próximos 12 meses?
- E nos primeiros 6 meses?

Distribuição Geométrica – para casa

- Todo final de semana você vai para a sua casa de campo. Você é meio apressado e gosta de ultrapassar o limite de velocidade na estrada. A probabilidade do radar pegar você acima da velocidade permitida é 15%.
- Se você é pego pela polícia tem que pagar uma multa de R$ 250,00 (por que, além de tudo você sempre esquece os documentos do carro em casa ....).

Distribuição Geométrica – para casa

- Suponha que cada ida para o campo no fim de semana seja uma repetição independente. O custo associado a cada viagem é R$ 25,00 (gasolina e pedágio).
- Você continua dirigindo em alta velocidade até receber a primeira multa.
- a) Qual o custo esperado deste procedimento (viajar em alta velocidade até ganhar a primeira multa)?
- b) Suponha que você tenha disponível R$ 300,00 no banco. Qual a probabilidade de você estourar o seu orçamento com este procedimento?