

Revisão de Estatística

Adaptada das aulas da Profa.
Jussara M. Almeida - UFMG

Por quê?

- Modelagem probabilística
- Avaliação dos resultados
 - Qual a probabilidade do tempo de residência no disco 1 ser inferior a 0.5 segundo?
 - Depende da distribuição de probabilidade do tempo de residência?
 - O tempo médio de resposta é uma boa estimativa do desempenho do sistema?
 - Depende da variabilidade de R: variância, desvio padrão
- Caracterização da carga
 - Como modelar o tempo entre chegada de requisições no servidor?

Variável Aleatória

- Uma variável aleatória (VA) X em um espaço amostral s é uma *função* $X: S \rightarrow \mathbb{R}$ que atribui um número real a cada ponto amostral em S

ou
- Uma variável aleatória é uma variável que recebe um valor numérico como resultado de um experimento.

Exemplos de Variável Aleatória

- Tempo entre chegadas de clientes em um servidor
- Número de visitas à CPU de uma requisição
- Tempo de residência no sistema
- Seja R_n o tempo de resposta de um cliente j que encontra n clientes na fila no momento de chegada
 - X_i = tempo de serviço de cliente i (iid)
 - Y = tempo residual do cliente em serviço quando o dado cliente j chega

Distribuição de Probabilidade de uma Variável Aleatória

- Função de distribuição acumulada (CDF) F da VA X é definida para todos os números reais b , $b \in \mathbb{R}$, como:

$$F(b) = P(X \leq b)$$

- Propriedades de uma CDF

F é uma função não decrescente: se $a < b \Rightarrow F(a) < F(b)$

$$\lim_{b \rightarrow \infty} F(b) = 1$$

$$\lim_{b \rightarrow -\infty} F(b) = 0$$

- CDF é uma caracterização completa de uma VA.

Variáveis Aleatórias Discretas

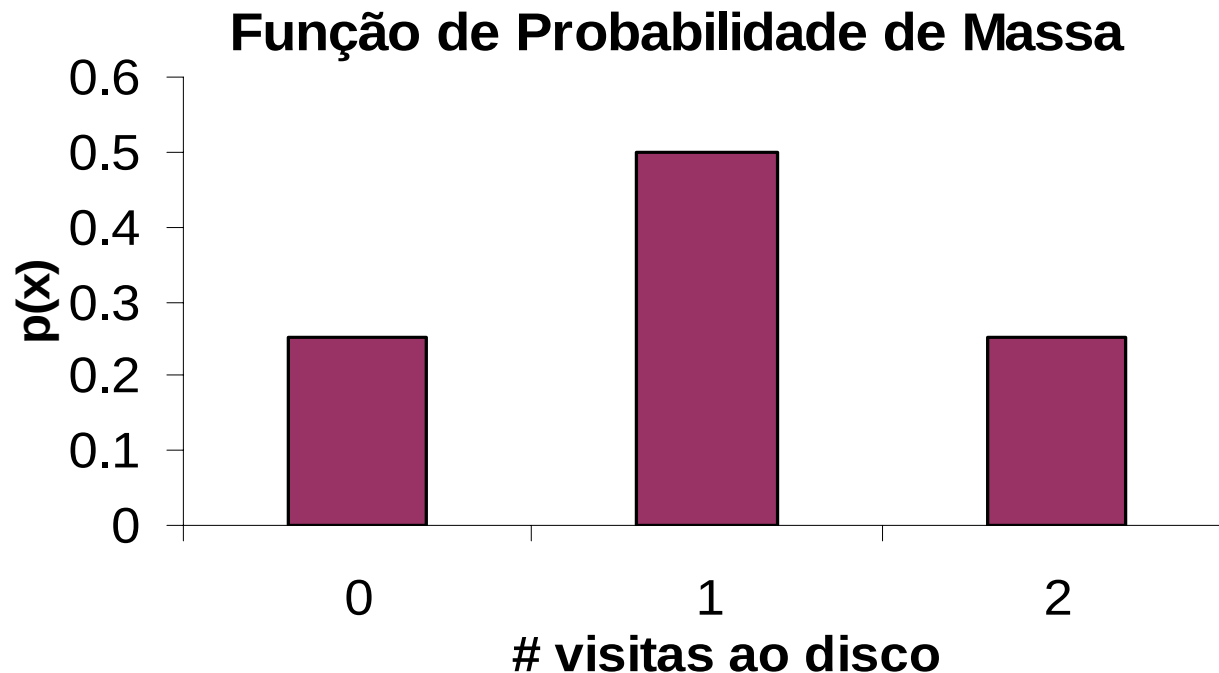
- Pode assumir um número contável de valores
- Função Probabilidade de Massa (PMF) $p()$:

$$p(a) = P[X = a]$$

- Propriedades:
 - $p(x_i) \geq 0, \quad i = 1, 2, \dots$
 - $p(x) = 0$, para todos outros valores de x
 $\sum_{i=1}^{\infty} p(x_i) = 1$
 - $F(x)$:
$$F(a) = \sum_{i=1}^a p(x_i)$$

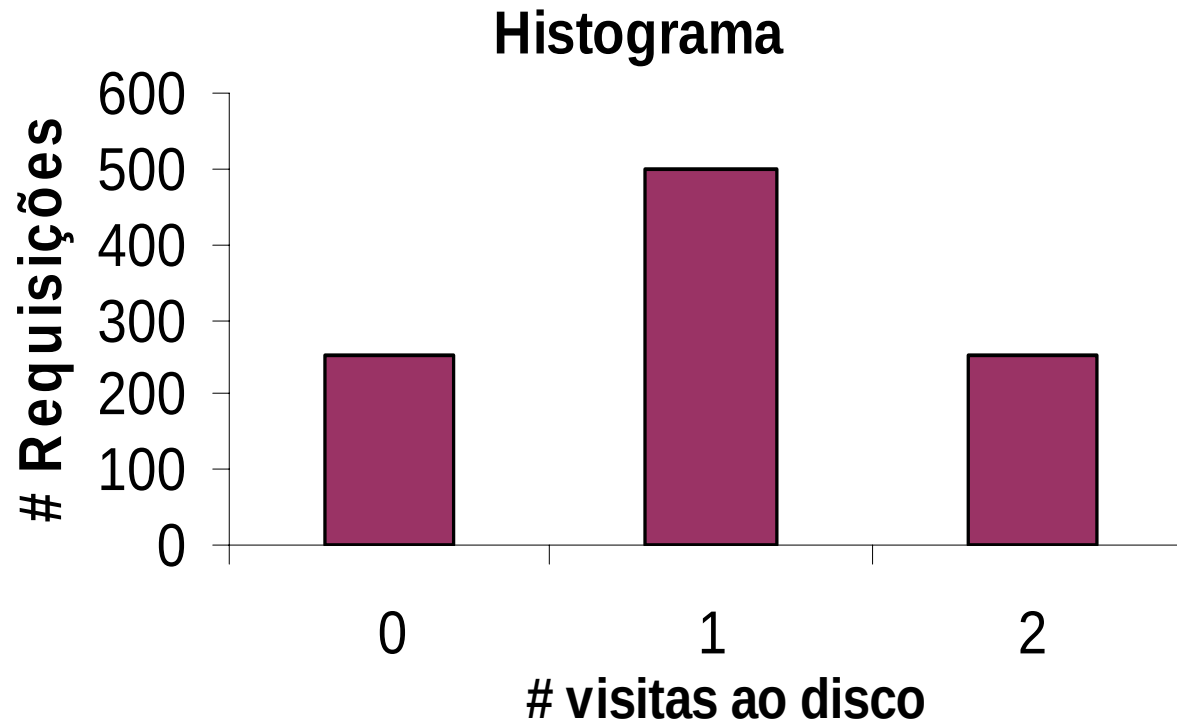
Função de Probabilidade de Massa

- Representação gráfica da $p(x)$:
 - Seja X o número de visitas que cada requisição faz ao disco
 - $p(X)$: $p(0) = 0.25$ $p(1) = 0.5$ $p(2) = 0.25$



Histograma

- Outra representação gráfica equivalente
 - Plota o número de vezes que a saída de um experimento aleatório foi igual a cada ponto amostral
 - Ex: se total de requisições ao servidor = 1000



Variáveis Aleatórias Contínuas

- Pode assumir um número incontável de valores
- Função Densidade de Probabilidade (f.d.p.):

$$P(a \leq X \leq b) = \int_a^b f(x)$$

- Propriedades:

- $f(x) \geq 0, \forall x$

- $\int_{-\infty}^{\infty} f(x) = 1$

- Função de Distribuição Acumulada: $F(a) = \int_{-\infty}^a f(x)$

Métricas Simples para Caracterização de uma Variável Aleatória

Expectativa / Esperanca

- Valor Esperado de uma VA X

$$E(X) = \sum_{x, p(x) > 0} xp(x)$$

X discreta

$$E(X) = \int_{-\infty}^{\infty} xf(x) dx$$

X contínua

- Ex: demanda média nos discos se:
 - 90% das requisições visitam disco 1,
 - tempo de serviço de disco 1 é 15 ms,
 - tempo de serviço do disco 2 é de 10 ms e
 - número de visitas por requisição é o mesmo para os dois discos e igual a 20 (arquivos com mesmo tamanho)

$$D = D_1 \times 0.9 + D_2 \times 0.1 = 20 \times 0.015 \times 0.9 + 20 \times 0.010 \times 0.1$$

Expectativa

- Expectativa de uma função da VA X

$$E(g(x)) = \sum_{x, p(x) > 0} g(x) p(x) \quad E(g(x)) = \int_{-\infty}^{\infty} g(x) f(x) dx$$

X discreta X contínua

- Ex: utilização média dos discos para $\lambda = 100$

$$U = \lambda D$$

$$U = U_1 \times 0.9 + U_2 \times 0.1$$

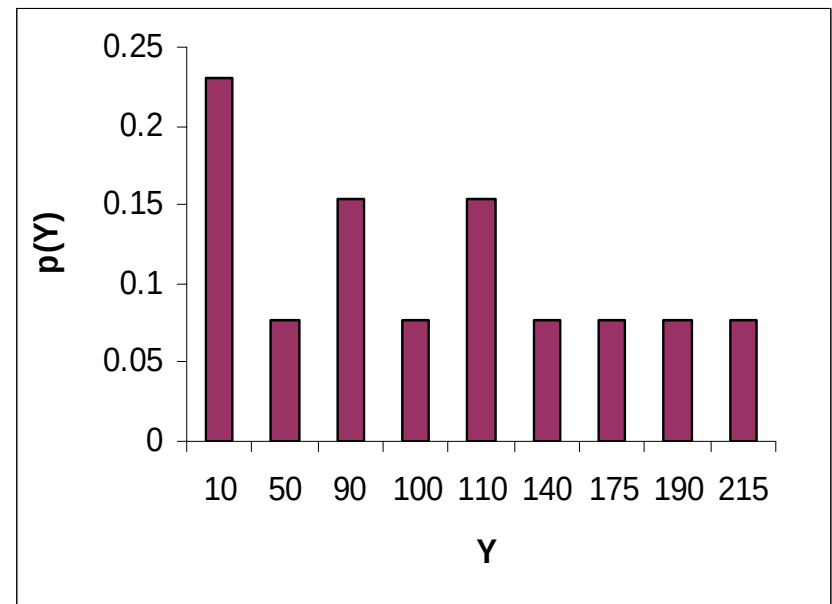
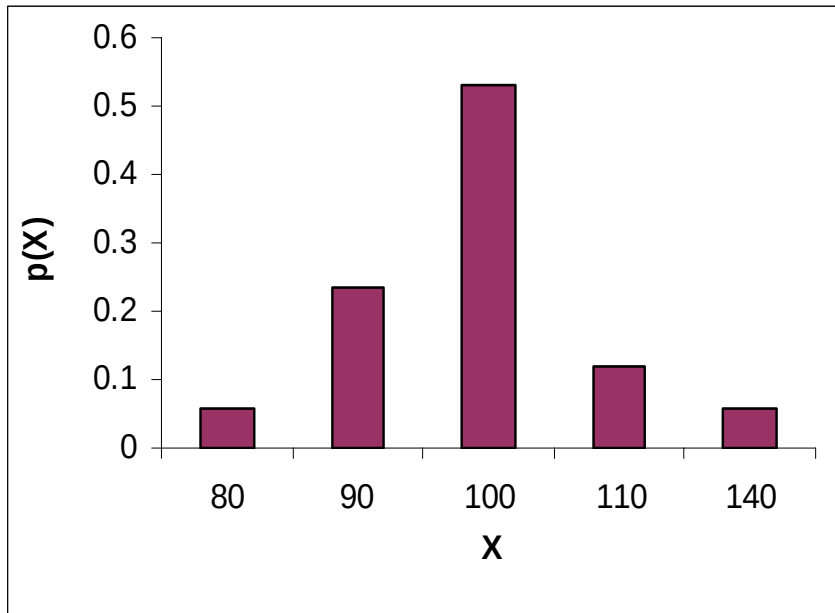
$$= \lambda \times D_1 \times 0.9 + \lambda D_2 \times 0.1 =$$

$$= 100 \times 20 \times 0.015 \times 0.9 + 100 \times 20 \times 0.010 \times 0.1$$

- k-esimo momento da VA X : $\overline{X^k} = E(X^k)$

Variância

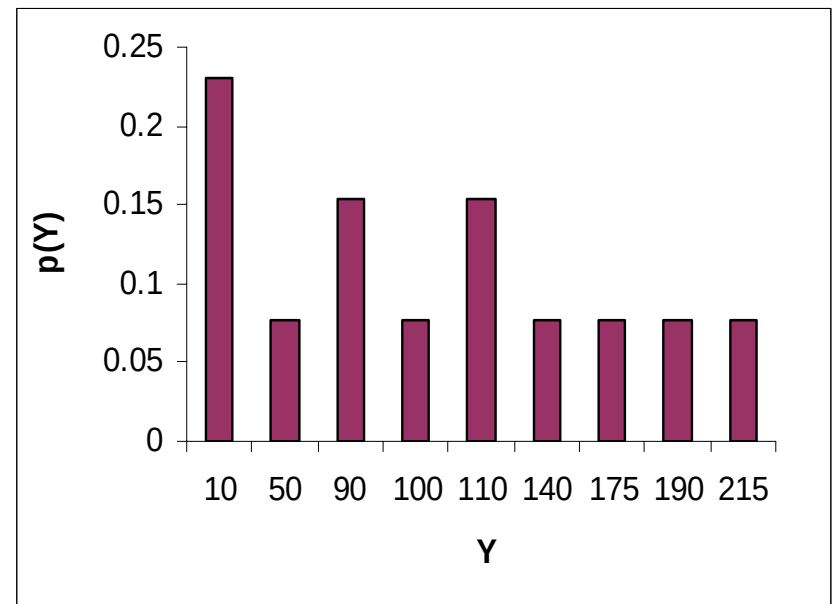
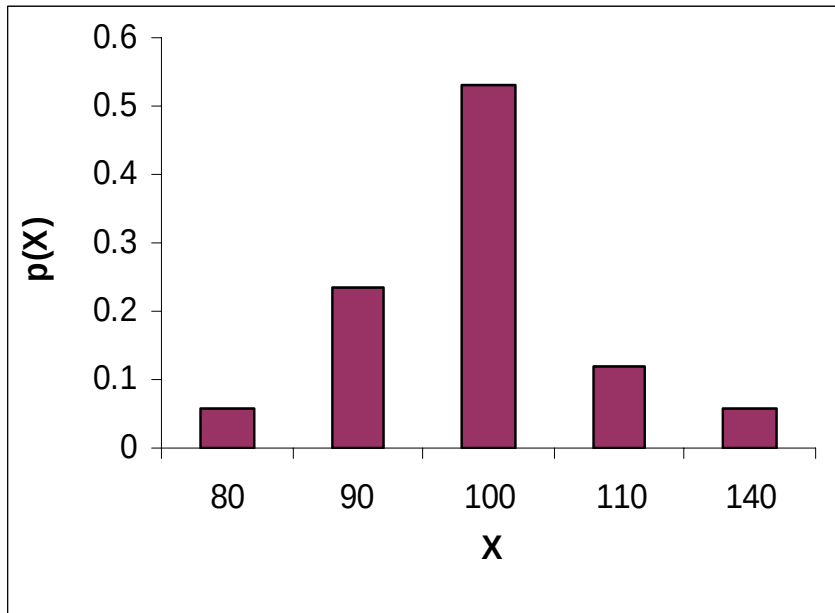
- Variância de uma VA X : variabilidade, espalhamento dos valores de X



$$E(X) = E(Y) = 100$$
$$\text{Var}(X) = 162.5 \quad \text{Var}(Y) = 4645.8$$

Variância

- Variância de uma VA X : variabilidade, espalhamento dos valores de X



$$\text{Var}(X) = E(X - \mu)^2 = E(X^2) - E(X)^2$$

$$\text{Var}(aX + b) = a^2 \text{Var}(X)$$

Desvio Padrão e Coeficiente de Variabilidade

- Desvio Padrão $SD(X)$

$$SD(X) = \sqrt{Var(X)}$$

- Coeficiente de Variacao (CV)

$$CV(X) = \frac{SD(X)}{E(X)}$$

- Diferentemente de SD , que depende da unidade das medições, o CV é uma medida sem unidade
- Mede a quantidade de variabilidade relativo ao valor médio
- Permite comparar a variabilidade existente em distribuições/amostras diferentes

Distribuição de Probabilidade

- Para uma caracterização completa de uma VA é preciso determinar a sua distribuição de probabilidade
 - $p(x)$ (se discreta) ou $f(x)$ (se contínua)
 - Existem várias distribuições discretas e contínuas na literatura que seguem comportamento bem definido
 - Conhecê-las é importante, pois podemos aplicar resultados previamente desenvolvidos
 - Ex: uma distribuição exponencial tem uma cauda mais leve do que uma distribuição Pareto
- Caudas pesadas têm impacto no consumo de recursos (ex: tempo de serviço)

Distribuições Discretas

- Bernoulli
- Binomial
- Poisson
- Geométrica
- Zipf
- Várias outras no livro do Ross

Distribuições Discretas

- Bernoulli (p)
 - $X = \{0,1\}$ $X = \{\text{sucesso}, \text{falha}\}$
 - $p(0) = P(X=0) = 1-p$
 - $p(1) = P(X=1) = p$
- Binomial (n, p)
 - $X = \#$ sucessos em n experimentos independentes, onde a probabilidade de sucesso em um experimento é p

$$p(i) = \binom{n}{i} p^i (1-p)^{n-i} \quad i = 0, 1, \dots, n$$

$$E(X) = np$$

$$\text{Var}(X) = np(1-p)$$

Distribuições Discretas

- Poisson (λ)
 - Número de eventos *independentes* que ocorrem em um intervalo de tempo (veja discussão em Ross, 4.8)
 - Número de chegadas em um servidor em 1 hora
 - Número de erros de impressão em uma página de um livro

$$p(i) = P(X = i) = e^{-\lambda} \frac{\lambda^i}{i!} \quad i = 0, 1, \dots, n$$

$$E(X) = \text{Var}(X) = \lambda$$

λ = # médio de eventos que ocorrem no período

- Aproximação para VA Binomial com n grande e p pequeno (Ross)
 - Se $X = \text{Binomial}(n, p)$, $X \approx \text{Poisson}(\lambda = np)$

Distribuições Discretas

- Poisson (λ)
 - Muito comumente usado para modelar chegada de sessões de usuários
 - servidores Web, multimídia, banco de dados, ftp, e-mail
 - Sessões são iniciadas por usuários
 - Chegada de duas sessões tendem a ser independentes: Poisson é uma boa aproximação
 - Contra-exemplo:
 - Chegada de *requisições* em um servidor Web
 - Premissa de independência não é válida: existe dependência entre requisições para o arquivo HTML e as imagens embutidas nele

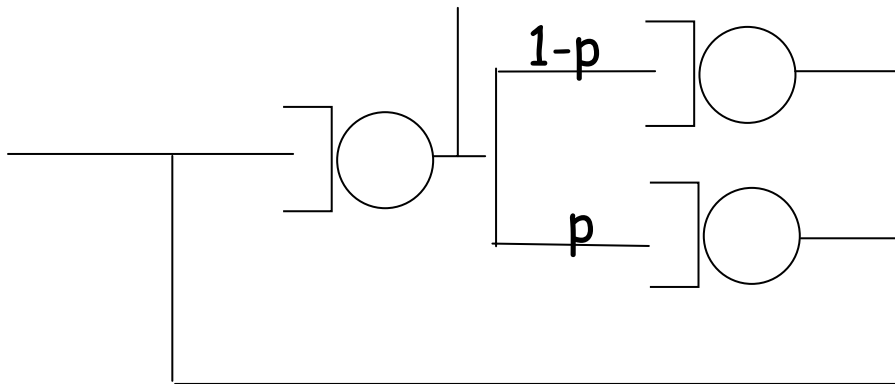
Distribuições Discretas

- Geométrica (p)

- Número de experimentos (sucesso/falha) até que um sucesso ocorra

$$p(n) = P(X = n) = (1 - p)^{n-1} p \quad n = 1, \dots, \infty$$

- Número de visitas ao disco 2



$$E(X) = 1/p$$
$$\text{Var}(X) = (1-p)/p^2$$

- Número de retransmissões de mensagem

Distribuições Discretas

- Zipf(α)
 - Comumente usada quando a distribuição é altamente concentrada em poucos valores
 - Popularidade de arquivos em servidores Web/multimídia
 - 90% dos acessos são para 10% dos arquivos
 - Popularidade de palavras na língua inglesa
 - Seja i , o elemento que ocupa a i -ésima posição no ranking de concentração

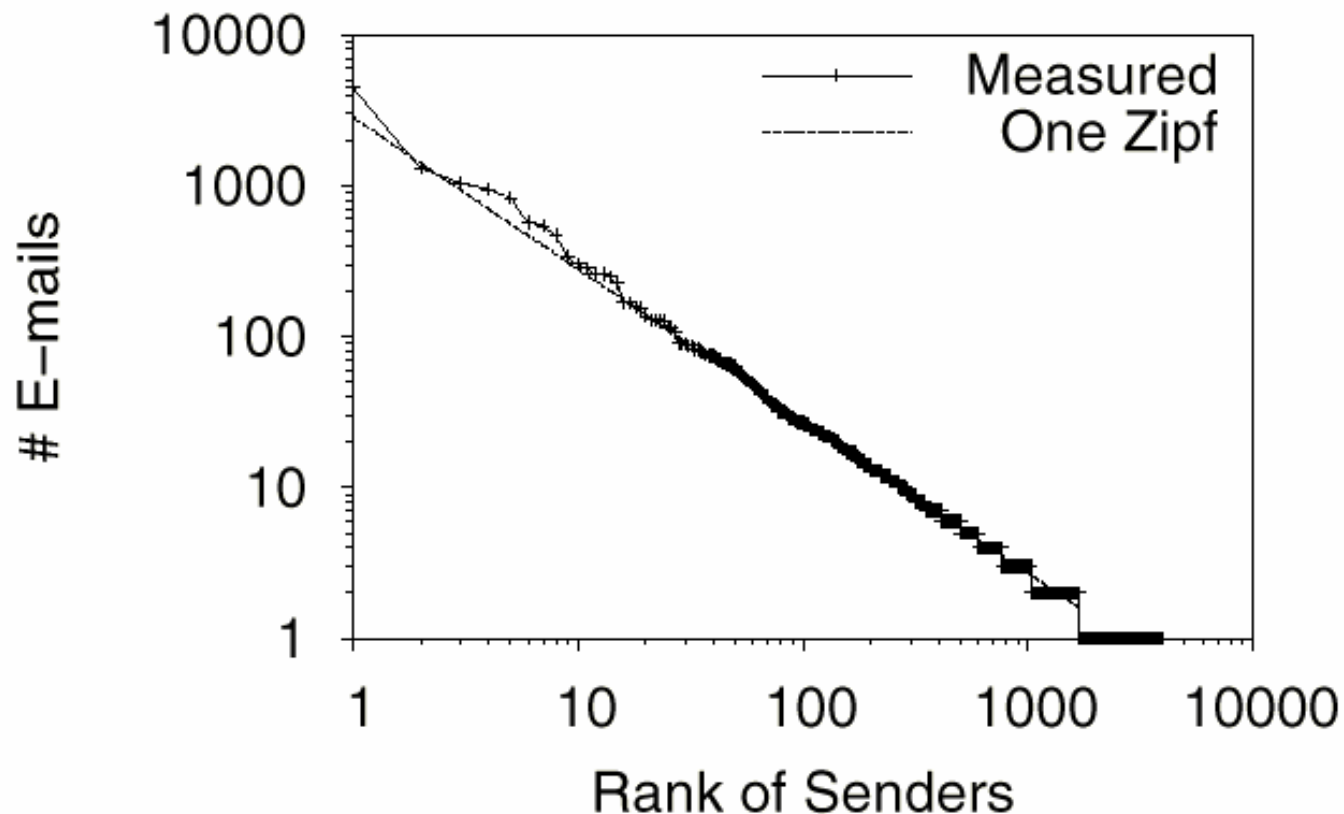
$$P(X = i) = \frac{C}{i^\alpha} \quad i = 1, 2, \dots$$

C é a constante de normalização

Zipf: lei das Potências

Distribuição Zipf

- Modela popularidade dos remetentes de e-mails para a UFMG



Distribuições Contínuas

- Uniforme
- Normal
- Exponencial
- Pareto
- LogNormal
- Gamma : ver Ross
- Weibull : ver Ross

Distribuições Contínuas

- Uniforme (a,b)

- X é uniformemente distribuída no intervalo [a,b], se

$$PDF : f(x) = \begin{cases} \frac{1}{b-a} & \text{se } a \leq x \leq b \\ 0 & \text{caso contrário} \end{cases}$$

$$CDF : F(x) = \begin{cases} 0 & \text{se } x < a \\ \frac{x-a}{b-a} & \text{se } a \leq x \leq b \\ 1 & \text{se } x \geq b \end{cases}$$

- Normal (μ , σ) ou Gaussiana

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

μ = valor esperado
 σ^2 = variância

Distribuições Contínuas

- Exponencial (λ)
 - Quantidade de tempo até que determinado evento ocorra
 - Tempo entre chegadas de sessões em um servidor

$$\begin{aligned} F(x) &= 1 - e^{-\lambda x} \\ f(x) &= \lambda e^{-\lambda x} \end{aligned} \quad x > 0 \quad \lambda > 0$$

λ = taxa de chegadas $1/\lambda$ = tempo médio entre chegadas

$$P(X \leq 1/\lambda) = 1 - e^{-\lambda \times 1/\lambda} = 1 - 1/e$$

$$E(X) = 1/\lambda$$

$$\text{Var}(X) = 1/\lambda^2 \Rightarrow \text{SD}(X) = 1/\lambda \Rightarrow \text{CV}(X) = 1$$

$$\text{CV} = 1 \not\Rightarrow \text{exponencial}$$

Distribuições Exponencial e Poisson

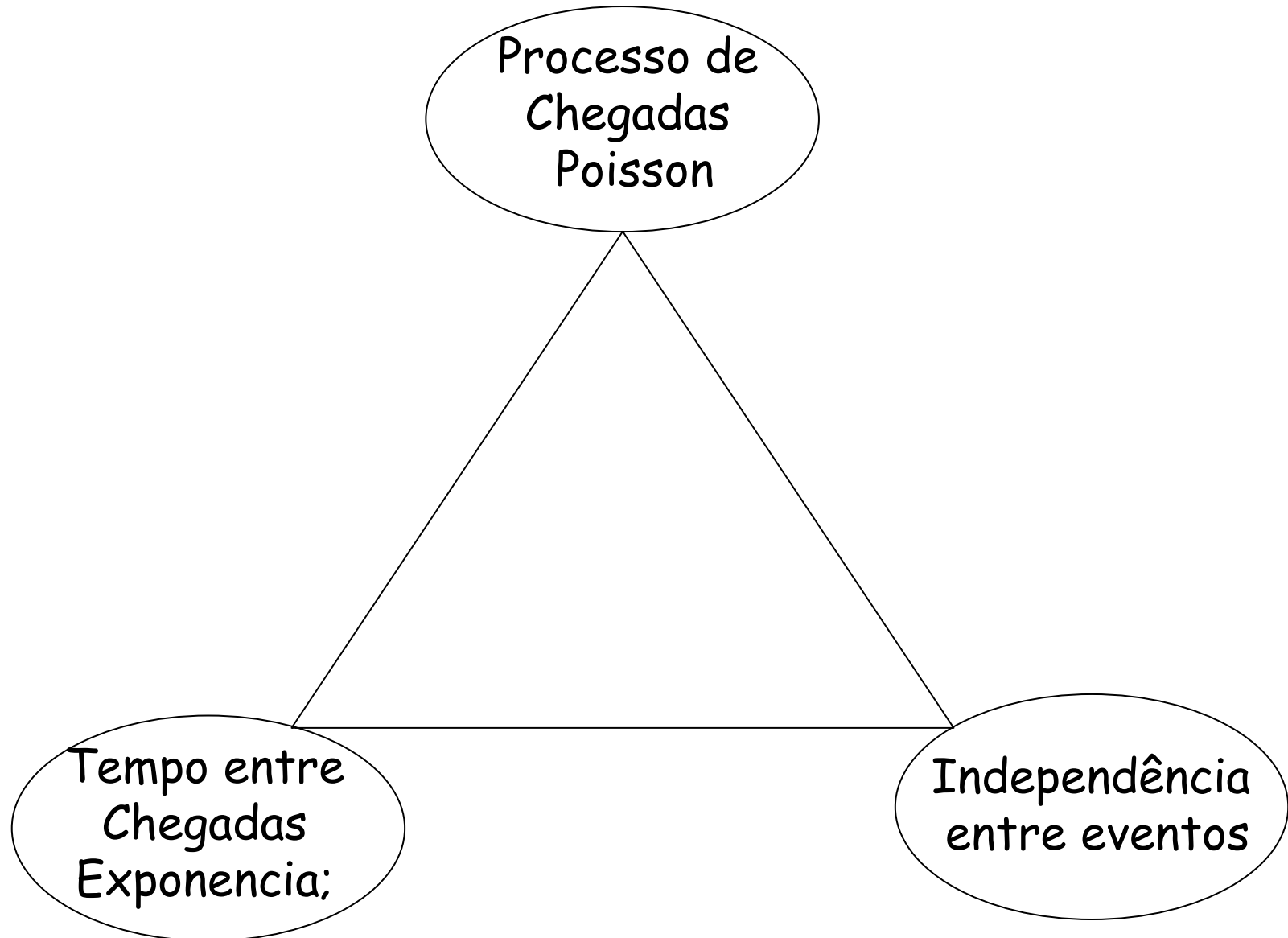
- Seja uma distribuição Poisson que denote o número de eventos $N(t)$ em um intervalo de tempo t
- Seja T_1 o momento do 1o evento
- Seja T_n o tempo entre o $(n-1)$ -esimo e o n -esimo eventos
- Sequência $\{T_n, n=1, 2, \dots\}$: tempos entre chegadas

$$P(T_1 > t) = P(N(t) = 0) = e^{-\lambda t} \Rightarrow T_1 \sim \text{exponencial}(\lambda)$$

$$\begin{aligned} P(T_2 > t \mid T_1 = s) &= \text{Prob (0 eventos em } (s, s+t) \mid T_1 = s) \\ &= \text{Prob (0 eventos em } (s, s+t) \text{ (eventos Poisson} \\ &\quad \text{são independentes))} \\ &= e^{-\lambda t} \Rightarrow T_2 \sim \text{exponencial}(\lambda) \end{aligned}$$

$\Rightarrow T_1, T_2, \dots, T_n$ são independentes e têm mesma distribuição exponencial(λ)

Distribuições Exponencial e Poisson



Distribuição Exponencial

- Exponencial (λ) :

$$P(X \leq t+x | X > t) = \frac{P([X \leq t+x] \cap [X > t])}{P(X > t)}$$

$$= \frac{P(t < X \leq t+x)}{1 - P(X \leq t)}$$

$$= \frac{P(X \leq t+x) - P(X > t)}{1 - P(X \leq t)}$$

$$= \frac{1 - e^{-\lambda(t+x)} - (1 - e^{-\lambda t})}{1 - (1 - e^{-\lambda t})}$$

$$= \frac{1 - e^{-\lambda t} e^{-\lambda x} - 1 + e^{-\lambda t}}{e^{-\lambda t}}$$

$$= \frac{e^{-\lambda t} (1 - e^{-\lambda x})}{e^{-\lambda t}} = 1 - e^{-\lambda x} = P(X \leq x)$$

Propriedade
sem memória
(memoryless)

Propriedade Memoryless

- Tempo de residência R de um cliente depende do # de clientes na fila quando ele chega ao centro, nos tempos de serviços destes clientes e no tempo que o cliente que está sendo servido no momento de chegada ainda permanecerá em serviço.
 - Seja X_i a VA para o tempo de serviço de cliente i na CPU
 - Seja X_i : exponencial(μ) para todos os clientes
 - Seja Y a VA que denota o tempo residual que o cliente que está em serviço no momento de chegada ainda permanecerá em serviço
 - Y também tem distribuição exponencial com parâmetro μ
 - Tempo que ainda falta independe do tempo que já esteve em serviço
- Nós usamos este resultado quando derivamos R para MVA
- Estado futuro não depende do estado passado

Propriedade Memoryless

- Distribuição **exponencial** é a única distribuição contínua que tem a **propriedade memoryless**
- Por sua vez, distribuição **geométrica** é a **única discreta** que tem a **propriedade memoryless**

Outras Distribuições Contínuas

- Pareto(α, k)

- Tamanho de arquivos Web
- Think time de um navegador Web
- Número de bytes em uma transferência FTP

$$f(X) = \alpha k^\alpha x^{-(\alpha+1)} \quad x \geq k \quad \alpha, k > 0 \qquad F(X) = 1 - \left(\frac{k}{x}\right)^\alpha \quad x \geq k$$

- Lognormal (μ, σ)

- Duração de sessões de usuários e de reqs interativas a vídeo
- Tamanho de e-mails
- Uma VA X é Lognormal (μ, σ) se $Y = \ln(X)$ é Normal (μ, σ)

$$f(X) = \frac{1}{\sqrt{2\pi}\sigma x} e^{-\frac{(\ln(x)-\mu)^2}{2\sigma^2}} \quad x \geq 0$$

Sumário

- Caracterização de uma VA X
 - Média de X
 - Variância, desvio padrão, CV
 - CDF
 - PMF (discreta) ou PDF (contínua)
 - Modelo de distribuição e seus parâmetros
 - Várias outras
- Melhor caracterização depende do que você está tentando calcular

Experimentando....

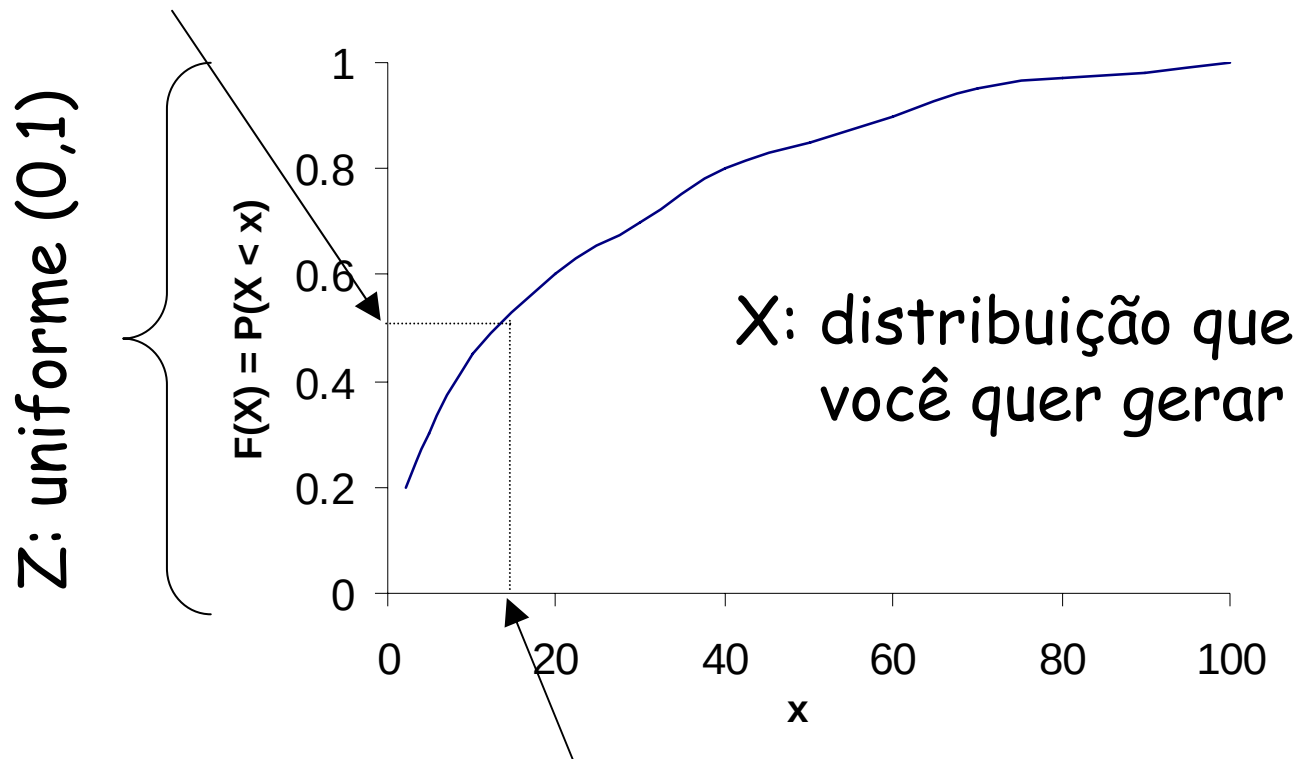
Gerando Distribuições

- Como gerar amostras de uma distribuição a partir de um gerador de números aleatórios uniformemente distribuídos (Unix: `random()`, `drand48()`)
 - Ex: geração de amostras de uma distribuição exponencial
$$F(X) = 1 - e^{-\lambda x}$$
$$Y = F^{-1}(X) = -1/\lambda \ln(1 - Z), \text{ onde } Z \sim \text{uniforme}(0,1) \text{ (} F(Z \leq z) = z \text{)}$$
$$F(Y) = P(Y \leq y) = P(-1/\lambda \ln(1 - Z) \leq y)$$
$$= P(\ln(1 - Z) \geq -\lambda y)$$
$$= P(1 - Z \geq e^{-\lambda y})$$
$$= P(Z \leq 1 - e^{-\lambda y}) = 1 - e^{-\lambda y} \Rightarrow Y \text{ é exponencial}$$

O mesmo procedimento pode ser utilizado para gerar amostras de diferentes distribuições, partindo da inversa da CDF da distribuição desejada

Gerando Distribuições

Gerador de números aleatórios
retorna valor entre 0 e 1. Ex: 0.52



Aplicando o número aleatório na função inversa de $F(X)$,
consegue-se gerar um ponto amostral

Métricas para Caracterização de uma Amostra de Dados

- Expectativa: Valor médio de um conjunto de amostras

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

- Variância de um conjunto de amostras s^2

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}$$

- Desvio Padrão s e CV calculados a partir de valor médio e s^2

Mediana, Quartil e Percentil

- Mediana:
 - Valor $\tilde{x}$ tal que 50% dos pontos estão abaixo dele
 - Ordene as observações em ordem crescente
 - $\tilde{x}$ = ao $(n+1)/2$ -esimo valor se n impar
 - $\tilde{x}$ = à média do $(n/2)$ -esimo e do $(n/2+1)$ -esimo valores, se n par
 - Divide as observações em duas partes
- Quartis:
 - Divide as observações em 4 partes
- Percentis:
 - Divide as observações em 100 partes
 - 99-esimo percentil separa as 1% maiores observações do restante

Combinando Variáveis Aleatórias

Funções de Distribuição Conjuntas

- CDF conjunta de X e Y

$$F_{X,Y}(a,b) = P(X \leq a \text{ e } Y \leq b) \quad X,Y: \text{ discretas ou contínuas}$$

- PMF conjunta de X e Y (X, Y discretas)

$$p_{X,Y}(a,b) = P(X = a \text{ e } Y = b)$$

- Probabilidade Condicional

$$p_{X|Y}(x|y) = p_{X,Y}(x,y) / p_Y(y) \quad (\text{discreta})$$

$$f_{X|Y}(x|y) = f_{X,Y}(x,y) / f_Y(y) \quad (\text{contínua})$$

- Expectativa

$$E[g(X,Y)] = \sum_x \sum_y g(x,y) p_{XY}(x,y)$$

$$E[g(X,Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) f_{XY}(x,y) dx dy$$

Funções de Distribuição Conjuntas

Se X e Y são independentes

$$F_{X,Y}(a,b) = F(a)F(b)$$

$$p_{X,Y}(a,b) = P(X=a)P(X=b)$$

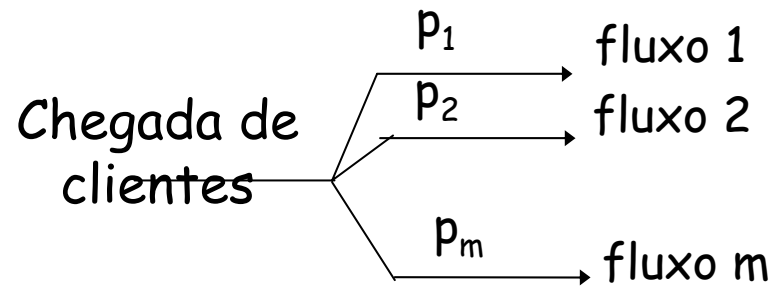
$$f_{X,Y}(x,y) = f(x)f(y) \quad (\text{contínuas})$$

$$\text{Prob}(X = a, Y \leq b) = p(X=a)F_{,Y}(b) \quad \begin{matrix} (X \text{ discreta,} \\ Y \text{ contínua}) \end{matrix}$$

$$E(g(X)h(Y)) = E(g(X)) \times E(h(Y))$$

Aplicações de Distribuição Conjunta

- Distribuição multinomial:



$$P(X_1 = n_1 \cap X_2 = n_2 \cap \dots \cap X_m = n_m) = \binom{n}{n_1 n_2 \dots n_m} p_1^{n_1} p_2^{n_2} \dots p_m^{n_m}$$

Aplicações de Distribuição Conjunta

- Transações no servidor A têm tempo de execução $T_A \sim \text{exponencial}(\lambda)$
- Transações no servidor B têm tempo de execução $T_B \sim \text{exponencial}(\mu)$
- Se duas transações T_1 e T_2 chegam no site A e B, respectivamente, ao mesmo tempo e são servidas imediatamente, qual a probabilidade de que T_1 termine antes que T_2

$$\begin{aligned} P(T_1 < T_2) &= \int_0^{\infty} P(T_1 \leq x \mid T_2 > x) P(T_2 > x) dx \\ &= \int_0^{\infty} P(T_1 \leq x) \times P(T_2 > x) dx = \int_0^{\infty} (1 - e^{-\lambda x}) e^{-\mu x} dx \\ &= \int_0^{\infty} e^{-\mu x} - e^{-(\lambda + \mu)x} dx = \left(-\frac{e^{-\mu x}}{\mu} + \frac{e^{-(\lambda + \mu)x}}{\lambda + \mu} \right) \Bigg|_0^{\infty} \\ &= \frac{\lambda}{\mu} \times \frac{1}{(\lambda + \mu)} \end{aligned}$$

Funções de Variáveis Aleatórias

Soma de Poissons

- X e Y são VAs independentes, $X \sim \text{Poisson}(\lambda_1)$ e $Y \sim \text{Poisson}(\lambda_2)$
- Qual a distribuição de $Z=X+Y$?

$$P(X + Y = n) = \sum_{k=0}^n P(X = k, Y = n - k) =$$

$$= \sum_{k=0}^n P(X = k)P(Y = n - k)$$

$$= \sum_{k=0}^n e^{-\lambda_1} \frac{\lambda_1^k}{k!} e^{-\lambda_2} \frac{\lambda_2^{n-k}}{(n-k)!}$$

$Z=X+Y$ é Poisson($\lambda_1 + \lambda_2$)

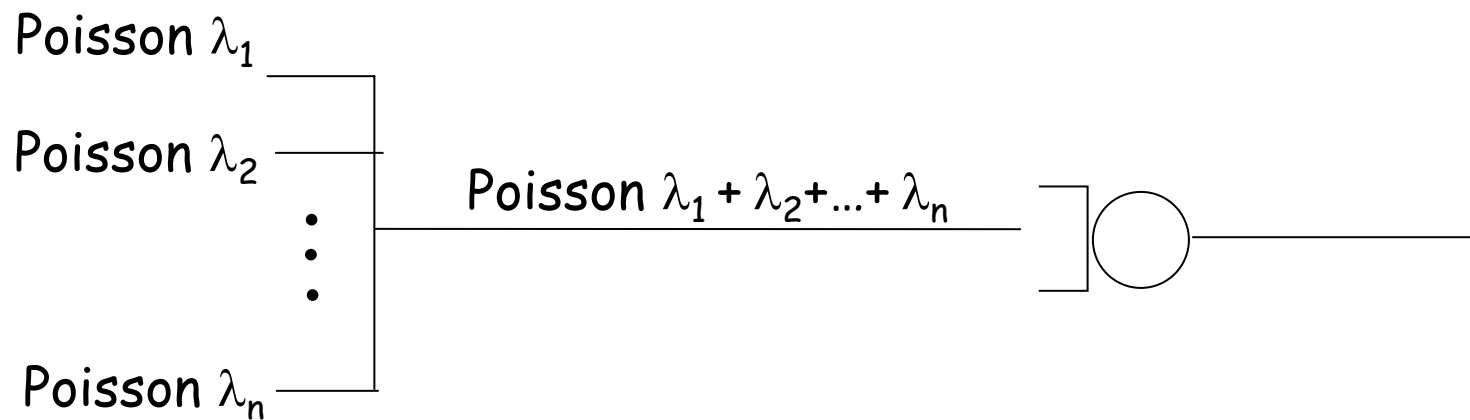
$$= e^{-(\lambda_1 + \lambda_2)} \sum_{k=0}^n \frac{\lambda_1^k \lambda_2^{n-k}}{k!(n-k)!}$$

$$= \frac{e^{-(\lambda_1 + \lambda_2)}}{n!} \sum_{k=0}^n \frac{n!}{k!(n-k)!} \lambda_1^k \lambda_2^{n-k} = \frac{e^{-(\lambda_1 + \lambda_2)}}{n!} (\lambda_1 + \lambda_2)^n$$

Teorema Binomial : $\sum_{i=0}^n \binom{n}{i} p^i q^{n-i} = (p + q)^n$

Aplicação: Soma de Poissons

- Chegada de sessões em um servidor Web, clientes estão em duas regiões. Mede separadamente em cada sessão e acha Poisson. Qual o processo de chegada agregado?



- Resultado se estende para n VAs independentes, cada uma com distribuição Poisson

Soma de Exponenciais

- X e Y são independentes, $X \sim \text{exponencial}(\lambda)$ e $Y \sim \text{exponencial}(\lambda)$
- Qual a distribuição de $Z=X+Y$?

$$F(Z = X + Y \leq z) = \int_0^z F(X \leq x | Y \leq z - x) F(Y \leq z - x) dx$$

$$= \int_0^z F(X \leq x) F(Y \leq z - x) dx$$

$$= \int_0^z \int_0^z f_X(x) f_Y(z - x) dx$$

$$= \int_0^z \int_0^z \lambda e^{-\lambda x} \times \lambda e^{-\lambda(z-x)} dx$$

$$= \int_0^z \lambda^2 e^{-\lambda z} \int_0^z dx = \int_0^z \lambda^2 z e^{-\lambda z} = \frac{\lambda^2 z^2}{2} e^{-\lambda z}$$

$Z=X+Y$ é Erlang com 2
estágios e parâmetro λ

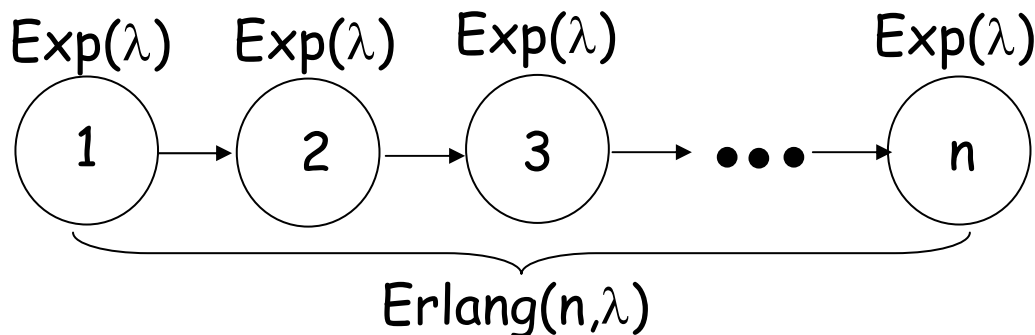
Aplicação: tempo de operação de um sistema com redundância
(X = tempo de operação de componente principal, Y = tempo do standby)

Soma de Exponenciais

- Genericamente: $X_1, X_2, \dots, X_n$, todas independentes e exponencial(λ): $Z = X_1 + X_2 + \dots + X_n \sim \text{Erlang de } n \text{ estágios}$

$$F(Z \leq z) = \sum_{k=0}^{n-1} \frac{(\lambda z)^k}{k!} e^{-\lambda z} \quad z \geq 0$$

- Ex: tempo de processamento dividido em várias etapas. A duração de cada etapa é exponencialmente distribuída com mesmo λ



- Se $X_i \sim \text{exponencial}(\lambda_i)$, onde λ_i são diferentes
 $Z = X_1 + X_2 + \dots + X_n \sim \text{Hipoexponencial (ver Trivedi)}$

Distribuição dos Mínimos

- Sistema composto de n componentes. Sistema funciona se ambos componentes estão operando corretamente
- Tempo de da falha : $X_1, X_2, \dots, X_n$ exponencial (λ)
- Tempo de de vida do sistema $Z = \min (X_1, X_2, \dots, X_n)$

$$P(Z \leq z) = P(\text{pelo menos um } X_i \leq z) = ?$$

$$P(\text{exatamente um } X_i \leq z) = ?$$

$$\begin{aligned} P(\text{exatamente } 1 \text{ } X_i \leq z) &= \binom{n}{1} F_X(z) (1 - F_X(z))^{n-1} \\ &= \binom{n}{1} (1 - e^{-\lambda z}) (1 - (1 - e^{-\lambda z}))^{n-1} \end{aligned}$$

Distribuição dos Mínimos

- $P(Z \leq z) = P(\text{pelo menos um } X_i \leq z)$

$$\begin{aligned} P(\text{pelo menos 1 } X_i \leq z) &= \sum_{j=1}^n \binom{n}{j} (F_X(z))^j (1 - F_X(z))^{n-j} \\ &= \sum_{j=1}^n \binom{n}{j} (1 - e^{-\lambda z})^j (e^{-\lambda z})^{n-j} \\ &= \sum_{j=1}^n \binom{n}{j} p^j (1 - p)^{n-j} \quad p = (1 - e^{-\lambda z}) \\ &= \sum_{j=0}^n \binom{n}{j} p^j (1 - p)^{n-j} - \binom{n}{0} p^0 (1 - p)^n \\ &= 1 - (1 - p)^n = 1 - \left(1 - (1 - e^{-\lambda z})\right)^n = 1 - e^{-\lambda n z} \end{aligned}$$

Z tem distribuição exponencial com parâmetro λn

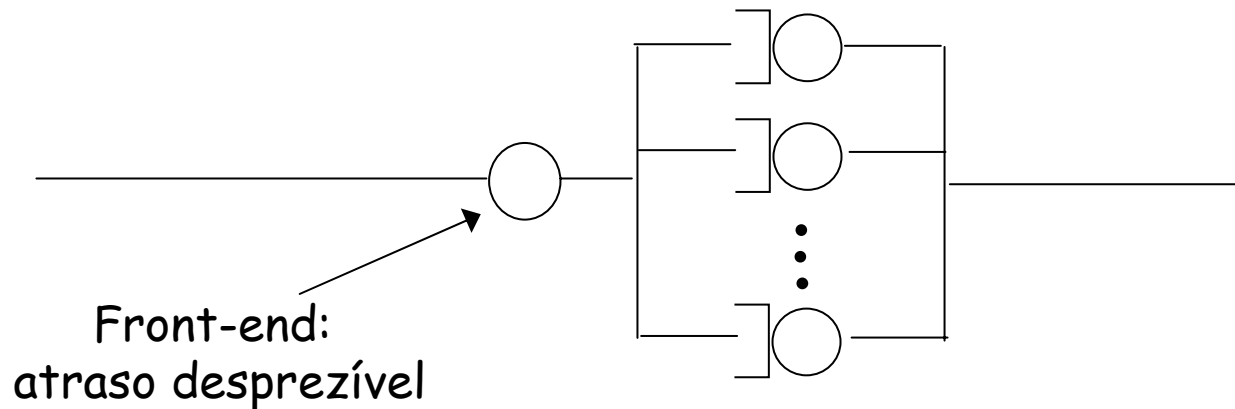
Generalizando: Estatística de ordem k é o k-esimo valor após ordenarmos as VAs em ordem crescente (Trivedi/Ross)

Distribuição dos Máximos

- n tarefas independentes : $X_1, X_2, \dots, X_n$: exponencial (λ)
- Tempo de resposta = tempo de execução da tarefa mais longa

$$Z = \max (X_1, X_2, \dots, X_n)$$

- Ex: tempo de resposta de máquina de busca composta de n processadores executando em paralelo. Cada máquina processa consulta em uma partição do dicionário



Distribuição dos Máximos

- n tarefas independentes : $X_1, X_2, \dots, X_n$: exponencial (λ)
- Tempo de resposta = tempo de execução da tarefa mais longa

$$Z = \max (X_1, X_2, \dots, X_n)$$

$$P(Z \leq z) = P(\max(X_i) \leq z)$$

$$= P(X_1 \leq z \cap X_2 \leq z \cap \dots \cap X_n \leq z)$$

$$= P(X_1 \leq z)P(X_2 \leq z) \dots P(X_n \leq z)$$

$$= (1 - e^{-\lambda z})(1 - e^{-\lambda z}) \dots (1 - e^{-\lambda z}) = (1 - e^{-\lambda z})^n$$

Exemplo

Considere um computador paralelo com n processadores. Sejam $X_1, X_2, \dots, X_n$, os tempos de falha dos processadores, cada um exponencialmente distribuído com parâmetro λ . Qual a distribuição da capacidade de processamento C_n do computador?

Ordene X_i 's em ordem crescente.

Seja Y_i a VA que ocupa a i -ésima posição: $Y_1 = \min(X_i)$

$$C_n = nY_1 + (n-1)(Y_2 - Y_1) + (n-2)(Y_3 - Y_2) + \dots + (n-j)(Y_{j+1} - Y_j) + \dots + (Y_n - Y_{n-1})$$

$$Y_1 = \min(X_i) \sim \text{exponencial}(n\lambda)$$

Exemplo (cont.)

Sejam $W_1, W_2, \dots, W_{n-j}$ os tempos restantes de processamento de cada um dos processadores ainda operando depois que j processadores falharam

$$Y_{j+1} - Y_j = \min(W_1, W_2, \dots, W_{n-j})$$

Pela propriedade memoryless da exponencial

$$W_i \sim \text{exponencial}(\lambda)$$

$$\text{Logo } Y_{j+1} - Y_j \sim \text{exponencial}((n-j) \lambda)$$

Exemplo (cont.)

Lembre que: $C_n = nY_1 + \dots + (n-j)(Y_{j+1} - Y_j) + \dots + (Y_n - Y_{n-1})$

Quais as distribuições de: $(n-j)(Y_{j+1} - Y_j)$ e nY_1

Se $X \sim \text{exponencial}(\lambda)$:

$$P(rX \leq x) = F(X \leq x/r) = 1 - e^{-\lambda/r x}$$

$$Y = rX \sim \text{exponencial}(\lambda/r)$$

$$nY_1 \sim \text{exponencial}(\lambda)$$

$$(n-j)(Y_{j+1} - Y_j) \sim \text{exponencial}(\lambda)$$

Logo $C_n =$ soma de $\text{exponencial}(\lambda)$: $C_n \sim \text{Erlang}(\lambda, n)$