Slides Semana 4

Função geradora de momentos (Bônus)

- Seja X uma variável aleatória discreta, com distribuição de probabilidade $p(x_i) = P(X = x_i)$, $i = 1, 2, \ldots$ A *função geradora de momentos* da variável X é dada por:

$$m_X(t) = \sum_{i=0}^{\infty} e^{tx_i} p(x_i).$$

- Seja X uma variável aleatória contínua com função de densidade $f(x)$. A *função geradora de momentos* da variável X é dada por:

$$m_X(t) = \mathrm{E}(e^{tx}) = \int_{-\infty}^{\infty} e^{tx} f(x) dx.$$

O domínio da função $m_X(t)$ consiste de todos os t para os quais e^{tX} tem valor esperado finito. Se $m_X(t)$ é finita em algum intervalo aberto que contenha a origem então tem-se que

$$m_X(t) = \mathrm{E}(e^{tX}) = \mathrm{E}\left(\sum_{n=0}^{\infty} \frac{t^n X^n}{n!}\right) = \sum_{n=0}^{\infty} \frac{\mathrm{E}(X^n)}{n!} t^n.$$

A expansão em serie de Taylor de $m_X(t)$ é

$$m_X(t) = \sum_{n=0}^{\infty} \frac{t^n}{n!} \frac{d^n}{dt^n} m_X(t)\Big|_{t=0}.$$

Comparando os coeficientes de t^n nas equações anteriores temos que os momentos ao redor da origem podem ser obtidos pelas sucessivas derivadas da função geradora de momentos avaliada em $t = 0$, isto é,

$$\mu_n' = \mathrm{E}X^n = \frac{d^n}{dt^n} m_X(t)\Big|_{t=0}.$$

Exemplo: Consideremos a seguinte distribuição de probabilidade

$p(x)$	0,3	0,5	0,2
x	1	2	3

Logo,

$$m_X(t) = \mathrm{E}(e^{tX}) = \sum_x e^{tx} p(x) = 0,3e^t + 0,5e^{2t} + 0,2e^{3t}.$$

Seja X uma v.a. com função de densidade dada por:

$$f_X(x) = \begin{cases} 2e^{-2x} & \text{se } x > 0, \\ 0 & \text{caso contrário} \end{cases}$$

e desta forma

$$m_X(t) = \mathrm{E}(e^{tX}) = \int_0^\infty e^{tx} 2e^{-2x} dx = \left(1 - \frac{t}{2}\right)^{-1} \text{ para } t < \frac{1}{2}.$$

Exemplo: Seja X uma variável aleatória discreta com distribuição binomial com parâmetros n e p , ou seja, $X \sim b(n,p)$. Então

$$\begin{aligned} m_X(t) = E\left(e^{tX}\right) &= \sum_{x=0}^{n} e^{tx} \binom{n}{x} p^x (1-p)^{n-x} \\ &= \sum_{x=0}^{n} \binom{n}{x} (e^t p)^x (1-p)^{n-x} = (pe^t + 1 - p)^n. \end{aligned}$$

Podemos encontrar a média e a variância através da função geradora de momentos. Assim

$$m'_X(t) = \frac{d}{dt}(pe^t + 1 - p)^n = n(pe^t + 1 - p)^{n-1}pe^t.$$

Então como

$$E(X) = m'_X(0) = n(p + 1 - p)^{n-1}p = np$$

. De forma, análoga

$$m''_X(t) = \frac{d^2}{dt^2}(pe^t + 1 - p)^n = n(n-1)(pe^t + 1 - p)^{n-2}pe^tpe^t + pe^tn(pe^t + 1 - p)^{n-1}$$

Logo

$$E\left(X^2\right) = m''_X(0) = n(n-1)p^2 + pn$$

, então obtemos que:

$$\mathrm{Var}(X) = E(X^2) - E(X)^2 = [n(n-1)p^2 + pn] - (np)^2 = np^2(n-1) + np - n^2p^2 = np(1-p).$$

Modelo Estatístico

Famílias paramétricas

A família $(\Omega, \mathcal{F}, \mathcal{P})$ é dita de modelo estatístico, i. é., temos Ω o espaço amostral como o conjunto dos possíveis valores que a variável aleatória X pode assumir, $\mathcal{F}$ a σ-'algebra de subconjuntos de Ω. A família

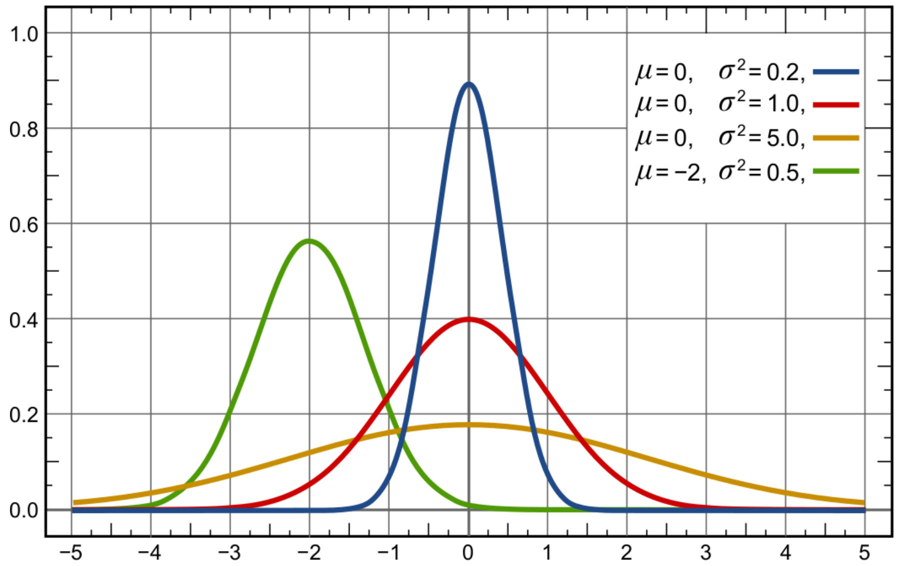
$$\mathcal{P} = \{P_\theta, \theta \in \Theta \subset \mathbb{R}^k\},\ k = 1, 2, \ldots .$$

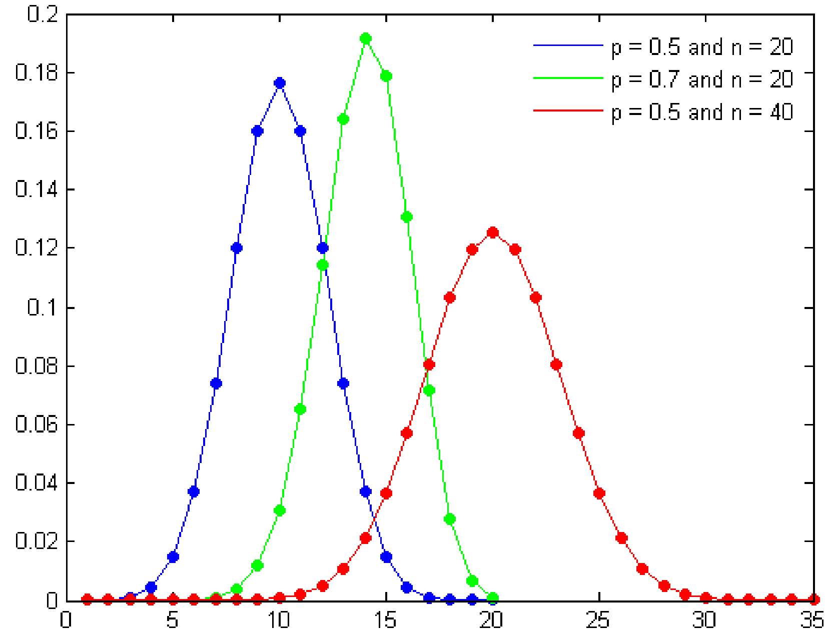
é uma família paramétrica de distribuições a qual a distribuição de X pertence. Aqui Θ é chamado de espaço paramétrico e corresponde ao conjunto dos possíveis valores para θ.

As distribuições P_θ pertencentes a $\mathcal{P}$ têm densidade $p_\theta(x)$ com respeito a uma mesma medida comum μ. Se as distribuições P_θ são absolutamente contínuas $p_\theta(x)$ é uma densidade com respeito a medida de Lesbegue - (comprimento de intervalos).

Se as distribuições P_θ são discretas $p_\theta(x)$ é uma densidade com respeito a uma medida de contagem - (contagem de elementos no conjunto).

Exemplos de famílias paramétricas





Distribuição Uniforme discreta

Seja X uma variável aleatória discreta sobre o conjunto $\{x_1, \ldots, x_N\}$ dizemos que X tem distribuição discreta uniforme e o denotamos por $X \sim \mathcal{U}\{x_1, \ldots, x_N\}$ se sua função de probabilidade está dada por:

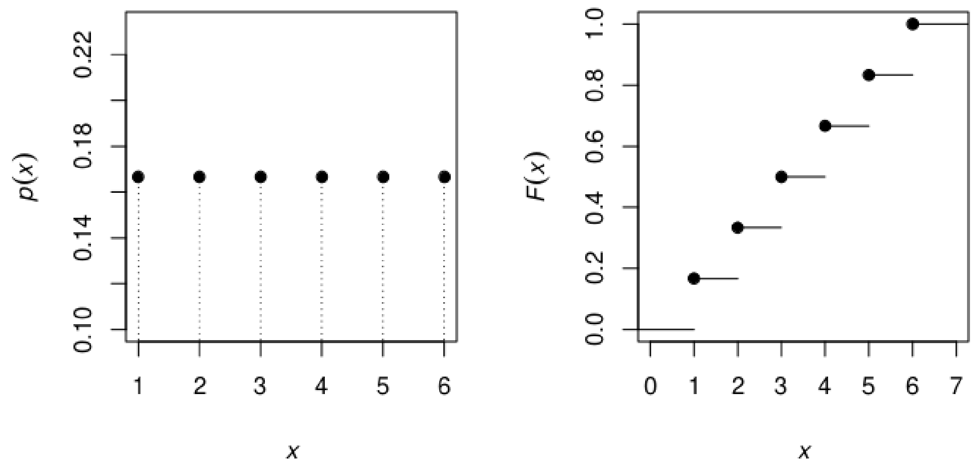
$$f(x) = f(x;N) = P(X = x) = \frac{1}{N}\mathbb{I}_{\{x_1,\ldots,x_N\}}(x) = \begin{cases} 1/N & \text{se } x = x_1, x_2, \ldots, x_N \\ 0 & \text{caso contrário} \end{cases}$$

Se o conjunto $\{x_1, \ldots, x_N\}$ é o conjunto dos inteiros de 1 até N então denotamos a distribuição discreta como $X \sim \mathcal{U}\{N\}$.

Teorema Se $X \sim \mathcal{U}\{x_1, \ldots, x_N\}$ então,

$$\begin{aligned} \mathrm{E}(X) &= \frac{1}{N}\sum_{i=1}^{N} x_i, \quad \text{(Valor esperado)} \\ \mathrm{Var}(X) &= \frac{1}{N}\sum_{i=1}^{N}(x_i - \mathrm{E}(X))^2, \quad \text{(Variância)} \\ m_X(t) &= \frac{1}{N}\sum_{i=1}^{N} e^{tx_i} \quad \text{(Função geradora de momentos)} \end{aligned}$$

Exemplo: Função de probabilidade e de distribuição uniforme sobre o conjunto $\{1, \ldots, 6\}$ é representada graficamente na Figura em que cada salto na função de distribuição de distribuição acumulada é 1/6.



Corolário: Se $X \sim \mathcal{U}\{N\}$ então

$$\begin{aligned}
\mathrm{E}(X) &= \frac{N+1}{2} \\
\mathrm{Var}(X) &= \frac{N^2-1}{12} \\
m_X(t) &= \frac{1}{N}\sum_{k=1}^{N} e^{tk} = \frac{e^t(1-e^{tN})}{N(1-e^t)}.
\end{aligned}$$

Corolário: Se $Y \sim \{a, a+k, a+2k, \ldots, b\}$ então

$$
\begin{aligned}
f(y) = P(Y=y) &= \frac{k}{b-a+k}. \\
m_Y(t) &= \frac{e^{at}(1-e^{ktN})}{N(1-e^{kt})}. \\
\mathrm{E}(Y) &= \frac{a+b}{2}. \\
\mathrm{Var}(Y) &= k^2\left(\frac{N^2-1}{12}\right).
\end{aligned}
$$

Nota: O conjunto $\{a, a+k, a+2k, \ldots, b\}$ é uma generalização do caso anterior no qual não requeremos que a sequência comece no valor 1 e os pontos sejam espaçados em um. Neste caso podemos observar que o tamanho do espaço amostral é

$$
N = \frac{b-a}{k} + 1 = \frac{b-a+k}{k}
$$

e assim a probabilidade para cada resultado é $f(y) = \dfrac{k}{b-a+k}$. Note que Y satisfaz a relação linear $Y = kX + (a-k)$ em que a variável aleatória $X \sim \mathcal{U}\{N\}$ onde $\{N\}$ representa o conjunto $\{1, 2, \ldots, N\}$.

Em geral a distribuição uniforme srve para modelar mecanismos de jogos (dados e moedas balanceados, cartas bem embaralhadas, etc).

Exemplo: Uma rifa tem 1000 bilhetes numerados de 1 a 1000. Tenho 5 bilhetes consecutivos numerados de 21 a 25, e meu colega tem outros 5 bilhetes, com os números 1, 11, 29, 68 e 93.

Quem tem maior probabilidade de ser sorteado?

- espalhar os números é a melhor forma de ganhar o sorteio?
- assumindo honestidade da rifa, todos os números têm a mesma probabilidade de ocorrência, com $\frac{1}{1000}$ para cada um.
- como eu e meu colega temos 5 bilhetes, temos a mesma probabilidade de ganhar a rifa: $\frac{5}{1000} = \frac{1}{200}$.
- assim, a probabilidade de ganhar depende somente da quantidade de bilhetes que se tem na mão, independente da numeração.

Ensaios de Bernoulli

Experimentos aleatórios em que apenas dois resultados são possíveis do tipo *sucesso* ou *fracasso*, *falha* ou *não falha*, dentre outros, são chamados *Ensaios de Bernoulli*. Tais experimentos apresentam resultados do tipo dicotômico. Denotamos por p como sendo a probabilidade de ocorrer *sucesso* e q a probabilidade de ocorrer o *fracasso*. Obviamente p e q devem ser não-negativos e $p + q = 1$.

O espaço amostral que representa um ensaio de Bernoulli é formado pelos dois possíveis resultados *sucesso* ou *fracasso*, ou seja: $\Omega = \{S, F\}$ em que $S =$ sucesso e $F =$ fracasso.

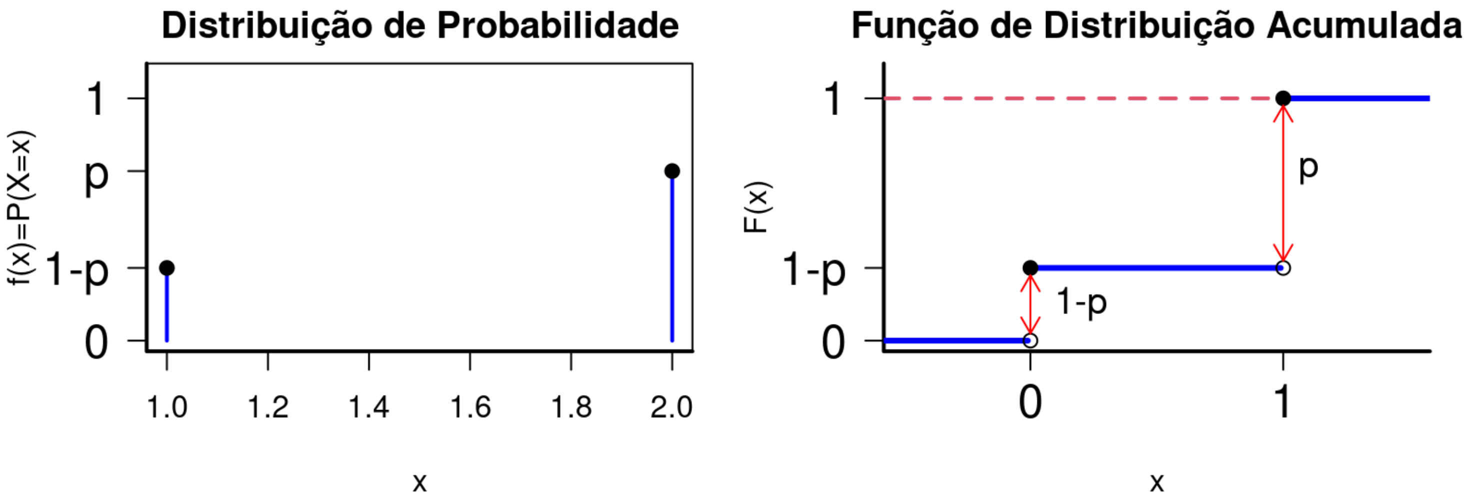
Definição: (Variável aleatória de Bernoulli) Para um ensaio de Bernoulli definimos a variável aleatória X como:

$$X = \begin{cases} 1 & \text{se ocorrer sucesso} \\ 0 & \text{em caso contrário.} \end{cases}$$

Assim

$$f(x) = f(x;p) = P(x = x) = p^x(1-p)^{1-x}\mathbb{I}_{\{0,1\}}(x) = f(x;p) = \begin{cases} p & \text{se } x = 1, \\ q & \text{se } x = 0, \\ 0 & \text{caso contrário.} \end{cases}$$

Aqui, $P(X = 1) = p$, e $P(X = 0) = q = 1 - p$.



Propriedades Se X segue uma distribuição de Bernoulli de parâmetro $0 \leq p \leq 1$ então:

1. $E(X) = p$
2. $Var(X) = p(1-p)$
3. $m_X(t) = p\mathrm{e}^t + (1-p)$

Nota: Se consideramos repetições independentes de n ensaios de Bernoulli, o espaço amostral cor-res-pon-den-te conterá 2^n pontos que serão combinações de sucessos e fracassos, ou seja:

$$\Omega = \{(S, S, S, \ldots, S), (F, S, S, \ldots, S), \ldots, (F, F, F, \ldots, F)\}.$$

Distribuição Binomial

Consideremos um experimento dicotômico em que os resultados possíveis sejam *sucesso* ou *fracasso*. Suponhamos que

$$P(\text{sucesso}) = p \qquad \text{e} \qquad P(\text{fracasso}) = 1 - p.$$

Consideremos n repetições independentes do experimento formando sequências de sucessos e fracassos dependendo do que tenha ocorrido.

Além disso suponhamos que p permanece constante durante as n realizações do experimento.

Seja X a variável aleatória que conta o número de sucessos nos n ensaios, então dizemos que X é uma variável aleatória com distribuição **binomial** com parâmetros n e p e a denotamos por $X \sim b(n;p)$.

Os valores possíveis de X são $0, 1, \ldots, n$ e a probabilidade de ocorrer um número k de sucessos é:

$$P(X = k) = \binom{n}{k} p^k q^{n-k} \qquad \text{para } k = 0, 1, \ldots, n.$$

Para mostrar que a forma das probabilidades binomiais são dadas pela $b(n;p)$ consideremos uma particular sequência de k sucessos. Isso implica na ocorrência de $n-k$ fracassos, ou seja:

$$\underbrace{SS\ldots S}_{k\text{ vezes}}\quad\underbrace{FFF\ldots F}_{n-k\text{ vezes}}$$

Como os ensaios são independentes, temos que as probabilidades se multiplicam, e a probabilidade de ocorrer k sucessos na sequência particular é dada por:

$$\underbrace{p.\,p.\ldots.\,p}_{k\text{ vezes}}\quad\underbrace{(1-p).\,(1-p).\ldots.\,(1-p)}_{n-k\text{ vezes}}=p^k.\,(1-p)^{n-k}.$$

Note que existem $\binom{n}{k}$ maneiras de obter k sucessos em n ensaios.

Se $X\sim b(n;p)$, então a probabilidade $P(X=k)$ como sendo a probabilidade de ocorrência de k sucessos define uma verdadeira distribuição de probabilidade. Isto é,

1. $P(X=k)\geq 0$
2. $\sum_{k=0}^{n}P(X=k)=1$

Para mostrar 1., basta notar que a probabilidade de ocorrência de k sucessos é uma multiplicação de valores positivos, pois $\binom{n}{k}\geq 0$, $p\geq 0$ e $q\geq 0$.

Para mostrar 2. temos que usar o a forma do binômio de Newton, i.e.

$$\sum_{k=0}^{n} P(X=k) = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} = (p+(1-p))^n = (p+1-p)^n = 1^n = 1.$$

O gráfico da distribuição binomial pode ser representado por um diagrama de barras. As barras representam as probabilidades de ocorrência do evento. Consideremos o caso em que $n = 20$ e $p = 0,1$, ou seja, $X \sim b(20; 0,1)$. Neste caso, as probabilidades assumem os seguintes valores:

$$P(X=0) = \binom{20}{0}(0,1)^0(1-0,1)^{20-0} = 0,12158$$

$$P(X=1) = \binom{20}{1}(0,1)^1(1-0,1)^{20-1} = 0,27017$$

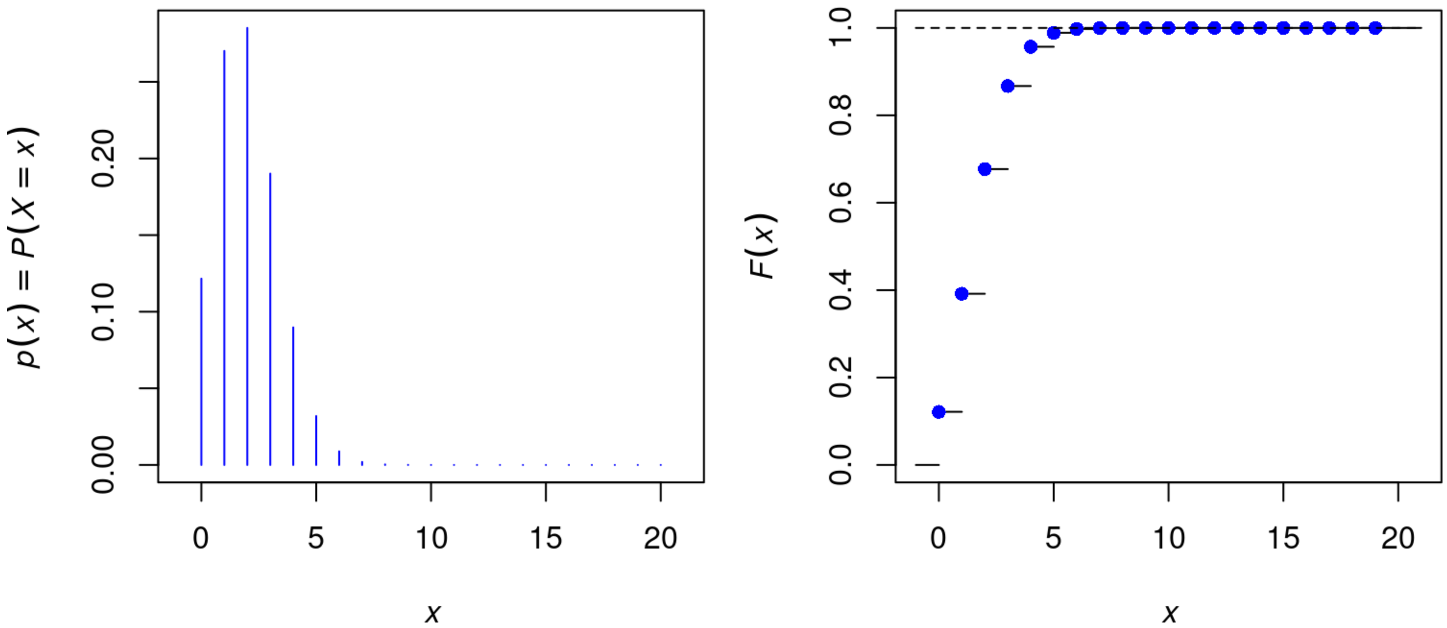
$$P(X=2) = \binom{20}{2}(0,1)^2(1-0,1)^{20-2} = 0,28518$$

$$\vdots$$

$$P(X=19) = \binom{20}{19}(0,1)^{19}(1-0,1)^{20-19} = 0,0000$$

$$P(X=20) = \binom{20}{20}(0,1)^{20}(1-0,1)^{20-20} = 0,0000$$

Note que:



Entre as possiveis aplicações da distribuição Binomila está a modelagem da quantidade de erros em um texto de n símbolos quando os erros entre símbolos são assumidos independentes e a probabilidade de erro em um símbolo do texto é igual a p; modelar o número de caras em n lançamentos de uma moeda que possui probabilidade p de cair cara em cada lançamento. Se $p = 1/2$, temos um modelo para o número de 1's em uma sequência binária de comprimento n escolhida aleatoriamente ou o número de caras em n lançamentos de uma moeda justa.

Valor esperado e variância da distribuição binomial: Quando $X \sim b(n;p)$, o valor esperado de X é dado por:

$$E(X) = \sum_{k=0}^{n} kp(k) = \sum_{k=0}^{n} k\binom{n}{k}p^k(1-p)^{n-k} = \sum_{k=0}^{n} k\frac{n!}{k!(n-k)!}p^k(1-p)^{n-k}$$
$$= \sum_{k=1}^{n} \frac{n!}{(k-1)!(n-k)!}p^k(1-p)^{n-k}.$$

Fazendo $k-1=r$ então $k=r+1$. Se $k=1$ temos que $r=0$ e se $k=n$ então $r=n-1$. Dai:

$$E(X) = \sum_{r=0}^{n-1} \frac{n!}{r!(n-(r+1)!}p^{r+1}(1-p)^{n-(r+1)}$$
$$= p\sum_{r=0}^{n-1} \frac{n(n-1)!}{r!((n-1)-r)!}p^r(1-p)^{(n-1)-r}$$
$$= np\sum_{r=0}^{n-1} \frac{(n-1)}{r!((n-1)-r)!}p^r(1-p)^{(n-1)-r}$$
$$= np\sum \binom{n-1}{r}p^r(1-p)^{(n-1)-r}.$$

Aplicando o Binômio de Newton temos que:

$$E(X) = np\binom{n-1}{r}p^r(1-p)^{(n-1)-r} = np[p+(1-p)]^{n-1} = np.$$

Seja X uma variável aleatória com distribuição Binomial (n, p). Então a função geradora de momentos de X, $m_X(t)$ é dada por $(pe^t + 1 - p)^n$. Logo $E(X) = np$ e $Var(X) = np(1-p)$.

Note que,

$$\begin{aligned}
m_X(t) &= E\left(e^{tX}\right) = \sum_{k=0}^{n} e^{tk}\binom{n}{k}p^k(1-p)^{n-k} \\
&= \sum_{k=0}^{n}\binom{n}{k}(e^t p)^k(1-p)^{n-k} = (pe^t+1-p)^n. \\
m'_X(t) &= \frac{d}{dt}(pe^t+1-p)^n = n(pe^t+1-p)^{n-1}pe^t \\
m''_X(t) &= \frac{d^2}{dt^2}(pe^t+1-p)^n = n(n-1)(pe^t+1-p)^{n-2}(pe^t)^2 + npe^t(pe^t+1-p)^{n-1}.
\end{aligned}$$

Logo de $E(X) = m'_X(0) = np$ e $E(X^2) = m''_X(0) = n(n-1)p^2 + pn$ obtemos o resultado.

Exemplo:

Exemplo (Comprador A ou B?): Um industrial fabrica peças, das quais 1/5 são defeituosas. Dois compradores **A** e **B** classificaram um grande lote de peças adquiridas em categorias I e II, pagando $\$1.20$ e $\$0.80$ por peça, respectivamente, do seguinte modo:

- **Comprador A**: retira uma amostra de cinco peças; se encontrar mais que uma defeituosa, classifica como II.
- **Comprador B**: retira uma amostra de dez peças; se encontrar mais que duas defeituosas, classifica como II.

Em média, qual comprador oferece mais lucro?

Fonte: Morettin & Bussab, Estatística Básica 5^a edição, pág 159.

Sabemos que 1/5 das peças são defeituosas. Podemos nos concentrar na probabilidade dos vendedores julgarem um lote como tipo I ou II.

Seja X o número de peças defeituosas em n testes.

O experimento do **comprador A** tem distribuição $X_A \sim Bin(5, 1/5)$ enquanto o experimento do **comprador B** tem distribuição $X_B \sim Bin(10, 1/5)$.

Para o **comprador A**, temos que:

$$
\begin{aligned}
P(X_A > 1) &= 1 - P(X_A = 0) - P(X_A = 1) \\
&= 1 - \binom{5}{0}\left(1 - \frac{1}{5}\right)^5 - \binom{5}{1}\left(\frac{1}{5}\right)\left(1 - \frac{1}{5}\right)^4 = 0.263
\end{aligned}
$$

De modo similar, para o **comprador B** temos:

$$
\begin{aligned}
P(X_B > 2) = 1 - \binom{10}{0}\left(1 - \frac{1}{5}\right)^{10} - \binom{10}{1}\frac{1}{5}\left(1 - \frac{1}{5}\right)^9 \\
- \binom{10}{2}\left(\frac{1}{5}\right)^2\left(1 - \frac{1}{5}\right)^8 = 0.322
\end{aligned}
$$

Como o segundo comprador irá classificar o lote como II com maior probabilidade que o primeiro, ele é o que oferece menor lucro para o fornecedor.

Mas podemos verificar o lucro esperado do vendedor. O preço por peça na categoria I: $1.20. Preço por peça na categoria II: $0.80.

Se o industrial decidir vender o lote para o **comprador A**, temos que

$$
E(\text{lucro A}) = 1.20 \times 0.737 + 0.80 \times 0.263 \approx 1.09
$$

Ou seja, ele irá lucrar em média $\$1.09$ por peça. Já se ele vender para o **comprador B**, temos que

$$E(\text{lucro B}) = 1.20 \times 0.678 + 0.80 \times 0.322 \approx 1.07$$

que é um lucro dois centavos inferior. Portanto, é mais interessante ao industrial que o comprador A examine mais peças.

Exemplo: Com base num estudo de consumo foi determinado que a preferência por dois tipos de marcas A e B de um dado produto encontram-se quase a par. Se as opções de compra entre estas marcas são independentes; Qual é a probabilidade de que entre 30 pessoas escolhidas ao acaso, não mais de dez tenham preferencia pela marca A?

Definimos X a variável aleatória que denota o número de pessoas que preferem a marca A. De acordo com a informação dada no problema temos que $X \sim b(30, 1/2)$. Portanto,

$$P(X \leq 10) = \sum_{k=0}^{10} \binom{30}{k} (1/2)^k (1/2)^{30-k} = 0,04936.$$

Distribuição de Poisson

Seja X uma variável aleatória discreta tomando valores inteiros positivos. Dizemos que X tem distribuição de Poisson com parâmetro λ, se sua função de probabilidade é

$$P(X=k)=\frac{e^{-\lambda}\lambda^k}{k!}, \quad \text{com } \lambda>0 \text{ e } k=0,1,2,\dots$$

Se X tem distribuição de Poisson de parâmetro λ a denotamos por $X \sim P(\lambda)$.

Se X for uma variável aleatória assumindo valores positivos cujas probabilidades são dadas por $P(\lambda)$, elas formam uma verdadeira distribuição de probabilidade, isto é: i) $P(X=k)>0$ e ii) $\sum_{k=0}^{\infty} P(X=k)=1$. A verificação de (i) é imediata, já que todas as quantidades envolvidas na $P(X=k)$ são positivas. Para verificarmos ii), temos que:

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

Podemos observar que

$$\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{\lambda},$$

que é o desenvolvimento em série de McLaurin da função $f(x) = e^x$. Assim,

$$\sum_{k=0}^{\infty} \frac{e^{-\lambda}\lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} e^{\lambda} = 1.$$

O gráfico da distribuição de Poisson pode ser representado por um diagrama de barras. Suponhamos por exemplo que $X \sim \lambda = 1$. Neste caso, a forma das probabilidades de Poisson, são:

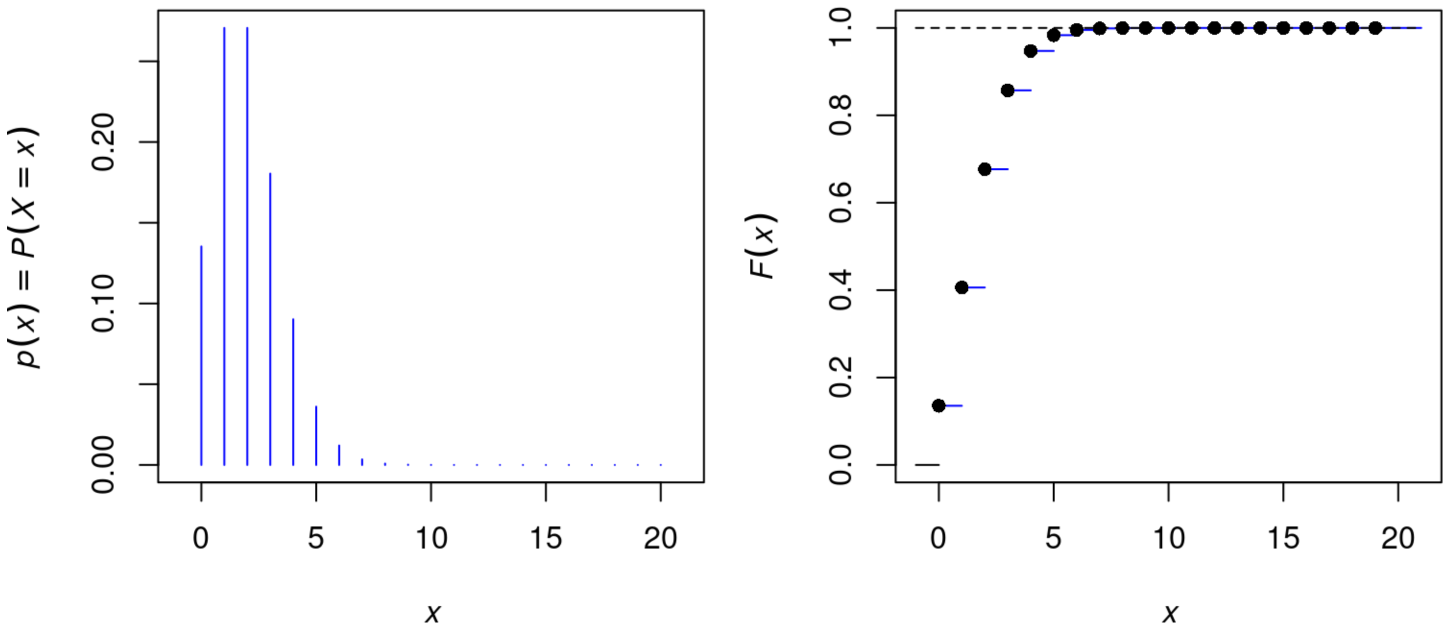
$$P(X = 0) = e^{-1}.1^0/0! = 0,367879$$

$$P(X = 1) = e^{-1}.1^1/1! = 0,367879$$

$$\vdots$$

$$P(X = 10) = e^{-1}.1^{10}/10! = 0,000000$$

e assim por diante.



Neste gráfico podemos observar que a distribuição de Poisson, é assimétrica a esquerda, porém se deslocando à direita a medida que λ aumenta.

Valor esperado e variância da distribuição de Poisson. Seja X uma variável aleatória com distribuição de Poisson com parâmetro λ. Então, o valor esperado de X é dado por:

$$E(X) = \sum_{k=0}^{\infty} kp(k) = \sum_{k=0}^{\infty} k\frac{e^{-\lambda}\lambda^k}{k!} = e^{-\lambda}\sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!}.$$

Fazendo $k-1=v$, então, $k=v+1$. Se $k-1=v$, então $v=0$ e, para $k\to\infty$, então $v\to\infty$.

$$E(X) = e^{-\lambda}\sum_{v=0}^{\infty} \frac{\lambda^{v+1}}{v!} = \lambda e^{-\lambda}\sum_{v=0}^{\infty} \frac{\lambda^v}{v!} = \lambda e^{-\lambda}e^{\lambda} = \lambda.$$

A variância de X é dada por:

$$Var(X) = E((X) - E(X))^2 = E(X^2) - (E(X))^2.$$

Note que,

$$E(X^2) = \sum_{k=0}^{\infty} k^2 p(k) = \sum_{k=0}^{\infty} k^2\frac{e^{-\lambda}\lambda^k}{k!} = \sum_{k=1}^{\infty} k\frac{e^{-\lambda}\lambda^k}{(k-1)!}.$$

Fazendo $k-1=v$, então, $k=v+1$, então $v=0$ e, para $k\to\infty$, então $v\to\infty$.

$$E(X^2)=\sum_{v=0}^{\infty}(v+1)\frac{e^{-\lambda}\lambda^{v+1}}{v!}=\lambda\left(\sum_{v=0}^{\infty}v\frac{e^{-\lambda}\lambda^{v}}{v!}+\sum_{v=0}^{\infty}\frac{e^{-\lambda}\lambda^{v}}{v!}\right)=\lambda(\lambda+1)=\lambda^2+\lambda.$$

Daí, obtemos o resultado

- Seja X uma variável aleatória com distribuição de Poisson de parâmetro λ. Então a função geradora de momentos é $m_X(t)=e^{\lambda(e^t-1)}$. Desta forma também pode ser deduzido que $E(X)=\lambda$ e $Var(X)=\lambda$. De fato, por definição

$$\begin{aligned} m_X(t) &= \mathrm{E}(e^{tX})=\sum_{x=0}^{\infty}e^{tx}P(X=x)=\sum_{x=0}^{\infty}e^{tx}\frac{\lambda^x e^{-\lambda}}{x!} \\ &= \sum_{x=0}^{\infty}e^{-\lambda}\frac{(\lambda e^t)^x}{x!}=e^{-\lambda}e^{\lambda e^t}=e^{\lambda(e^t-1)}. \end{aligned}$$

Assim,

$$E(X)=\frac{d}{dt}m_X(t)\Big|_{t=0}=e^{\lambda(e^t-1)}\lambda e^t\Big|_{t=0}=\lambda.$$

Agora,

$$Var(X) = \frac{d^2}{dt^2} m_X(t)\Big|_{t=0} - (E(X))^2 = \left[\lambda e^t e^{\lambda(e^t-1)} + \lambda^2 e^{2t}\, e^{\lambda(e^t-1)}\right]\Big|_{t=0} - \lambda^2$$
$$= \lambda e^t e^{\lambda(e^t-1)} \left[1 + \lambda e^t\right] \big|_{t=0} - \lambda^2 \; = \lambda(\lambda+1) - \lambda^2 = \lambda.$$

Teorema Seja X uma variável aleatória com distribuição binomial de parâmetro n e p.
Quando $n \to \infty$ e $np = \lambda$ (constante), ou seja, quando $n \to \infty$ e $p \to 0$ de modo que $np = \lambda$, temos que

$$\lim_{n\to\infty} P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!},$$

que é a aproximação da distribuição binomial a uma distribuição de Poisson.

Para exemplificar consideremos a seguinte tabela que apresenta a boa ``qualidade” da aproximação da distribuição Binomial para $P(X = k)$ da distribuição de Poisson
quando $\lambda = np = 1$.

k	Poisson	$b(100, \frac{1}{100})$	$b(10, \frac{1}{10})$
0	0,367	0,366	0,349
1	0,367	0,369	0,387
2	0,184	0,184	0,194
3	0,061	0,061	0,057

Aplicações da distribuição de Poisson: A opção pela utilização da distribuição de Poisson em um determinado experimento exige algumas observações:

1. O número de ocorrência do evento em qualquer intervalo de tempo depende somente da duração do intervalo de tempo, ou seja, quanto maior o intervalo, maior tende a ser o número de ocorrências do evento;

2. O eventos ocorrem independentemente, ou seja, um excesso ou falta de ocorrência do evento num determinado intervalo de tempo não interfere sobre a ocorrência de eventos ocorridos durante qualquer outro intervalo;

3. A possibilidade de ocorrência de dois ou mais eventos num pequeno intervalo de tempo é muito pequena, ao ser comparado com uma única ocorrência do evento.

Exemplo (Magalhães e Lima, 2002). A emissão de partículas radioativas tem sido modelada através de uma distribuição de Poisson, com o valor do parâmetro dependendo da fonte utilizada. Suponha que o número de partículas alfa emitidas por minuto, seja uma variável aleatória seguindo o modelo de Poisson com parâmetro 5, isto é, a taxa média de ocorrência é de 5 emissões a cada minuto. Qual é a probabilidade de haver mais de 2 emissões em um minuto? Podemos supor que $X \sim P(5)$. Assim:

$$P(X > 2) = 1 - [P(X = 0) + P(X = 1) + P(X = 2)]$$

$$P(X = 0) = \frac{e^{-5}.5^0}{0!} = 0,006738,$$

$$P(X = 1) = \frac{e^{-5}.5^1}{1!} = 0,03369,$$

$$P(X = 2) = \frac{e^{-5}.5^2}{2!} = 0,08424.$$

Logo, a probabilidade de emissão de mais de 2 partículas é: $P(X > 2) = 1 - 0,124669 = 0,875331.$

Distribuição Normal

Distribuição Normal

Dizemos que uma v.a. X possui distribuição normal (distribuição Gaussiana) com parâmetros μ e σ^2, $\mu \in \mathbb{R}$ e $\sigma^2 > 0$, se a f.d.p. f_X é dada por:

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right], \qquad -\infty < x < \infty.$$

Notação: $X \sim N(\mu, \sigma^2)$.

A esperança e variância de uma v.a. $X \sim N(\mu, \sigma^2)$ são: $E(X) = \mu$ e $Var(X) = \sigma^2$.

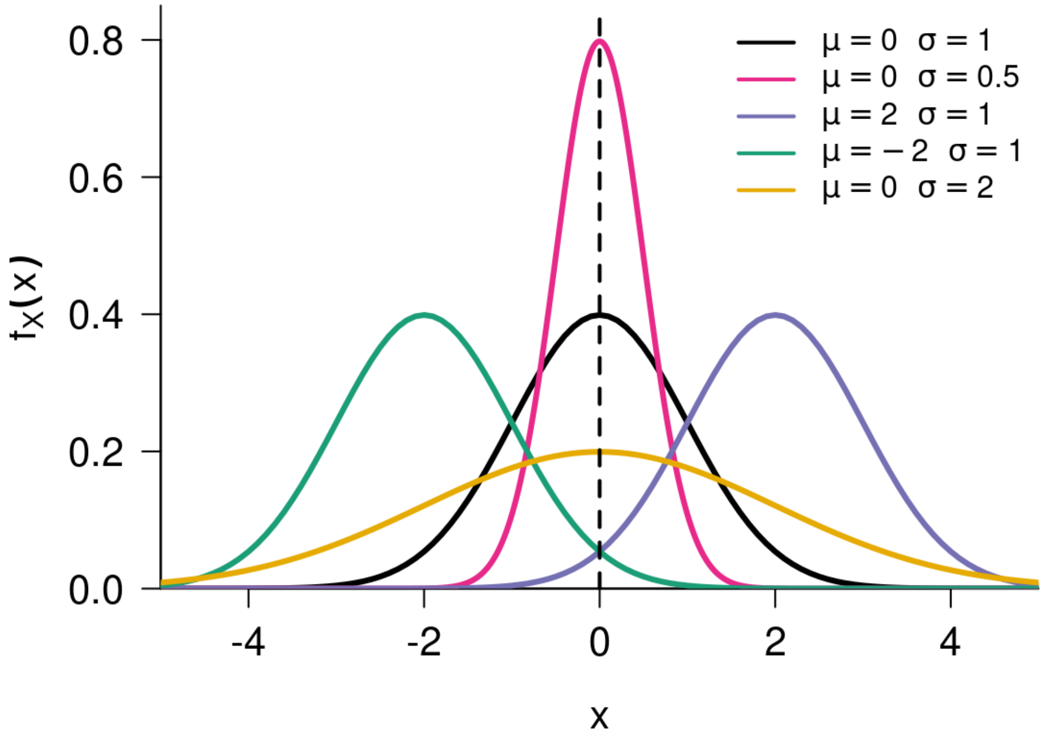
Esperança:

$$E(X) = \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] dx = \mu.$$

Variância:

$$Var(X) = \mathbb{E}([X - \mathbb{E}(X)]^2) = \int_{-\infty}^{\infty} (x-\mu)^2 \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] dx = \sigma^2.$$

Representação gráfica da função de densidade de probabilidade de uma v.a. $X \sim N(\mu, \sigma^2)$:



A **função de densidade** apresenta a “Forma de sino”, centrada em μ e escala (dispersão) controlada por σ^2.

Propriedade: Se $X \sim N(\mu, \sigma^2)$, então a transformação

$$Z = \frac{X - \mu}{\sigma} \sim N(0, 1).$$

Dizemos que Z tem distribuição **Normal Padrão** e sua densidade se reduz a:

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, \qquad -\infty < z < \infty.$$

A função de distribuiiçã acumulada (f.d.a) de uma Normal padrão, que denotaremos por Φ, é:

$$\Phi(z) = P(Z \leq z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt.$$

Gráficos da distribuição normal padrão

- $\phi(x)$ é contínua em todo $\mathbb{R}$.
- $\phi(x)$ é simétrica com respeito do eixo das ordenadas, isto é $\phi(t) = \phi(-t)$.
- A função $\phi(x)$ nunca assume o valor zero, mas tem como assíntota horizontal o eixo das abscissas, isto é

$$\lim_{x \to -\infty} \phi(x) = \lim_{x \to \infty} \phi(x) = 0.$$

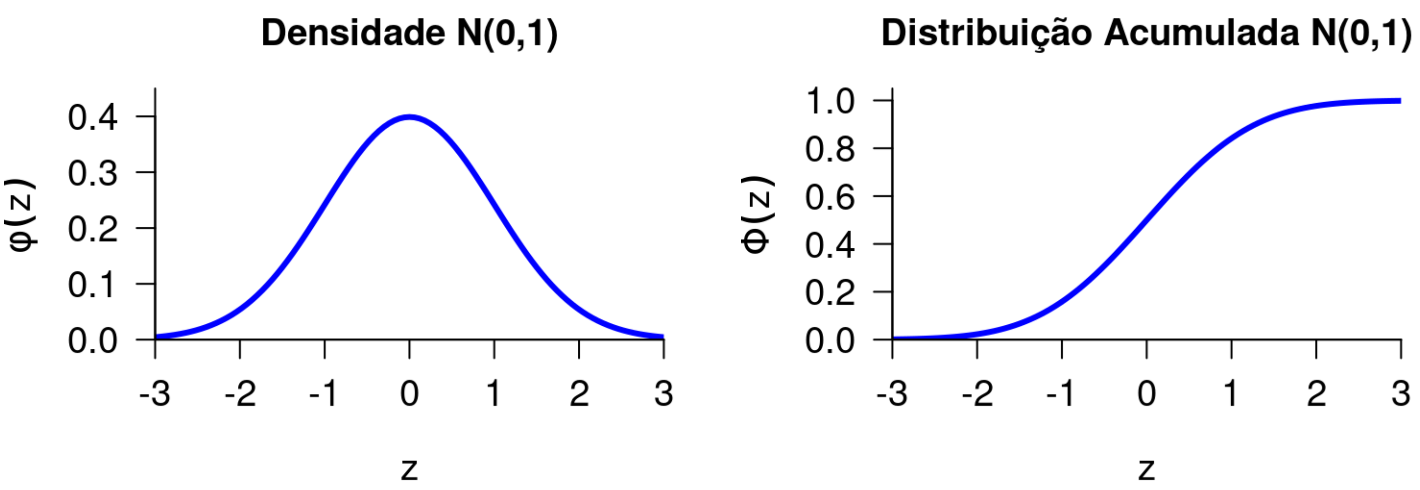
- A função $\phi(x)$ satisfaz:

$$\phi'(x) = -x\phi(x) = \begin{cases} > 0 & \text{se } x < 0, \text{ ou seja } \phi(x) \text{ é crescente,} \\ < 0 & \text{se } x > 0, \text{ ou seja } \phi(x) \text{ é decrescente.} \end{cases}$$

- $\phi(x)$ tem um ponto de máximo em 0 que é igual a $\frac{1}{\sqrt{2\pi}}$. De fato, $\phi'(x) = 0$, se, e somente se, $x = 0$ e $\phi''(x) = -\phi(x) + x^2\phi(x)$ o que implica em $\phi''(0) = -\phi(0) < 0$.

- A função $\phi(x)$ apresenta dois pontos de inflexão, um $x = -1$ e outro em $x = 1$. A função é concava para abaixo se $-1 < x < 1$ e concava para acima em caso contrário.

$$\phi''(x) = (-1 + x^2)\phi(x) = 0 \text{ implica } x^2 = 1 \text{ o que implica } x = \pm 1.$$



Exemplo: SAT e ACT

Uma universidade americana recebeu inscrição de dois alunos (Pam e Jim) com os respectivos históricos escolares. No entanto, Pam realizou o SAT e tirou 1800, enquanto que o Jim fez o ACT e tirou 24. **Como a universidade pode comparar os dois alunos, baseando-se nesses testes?**

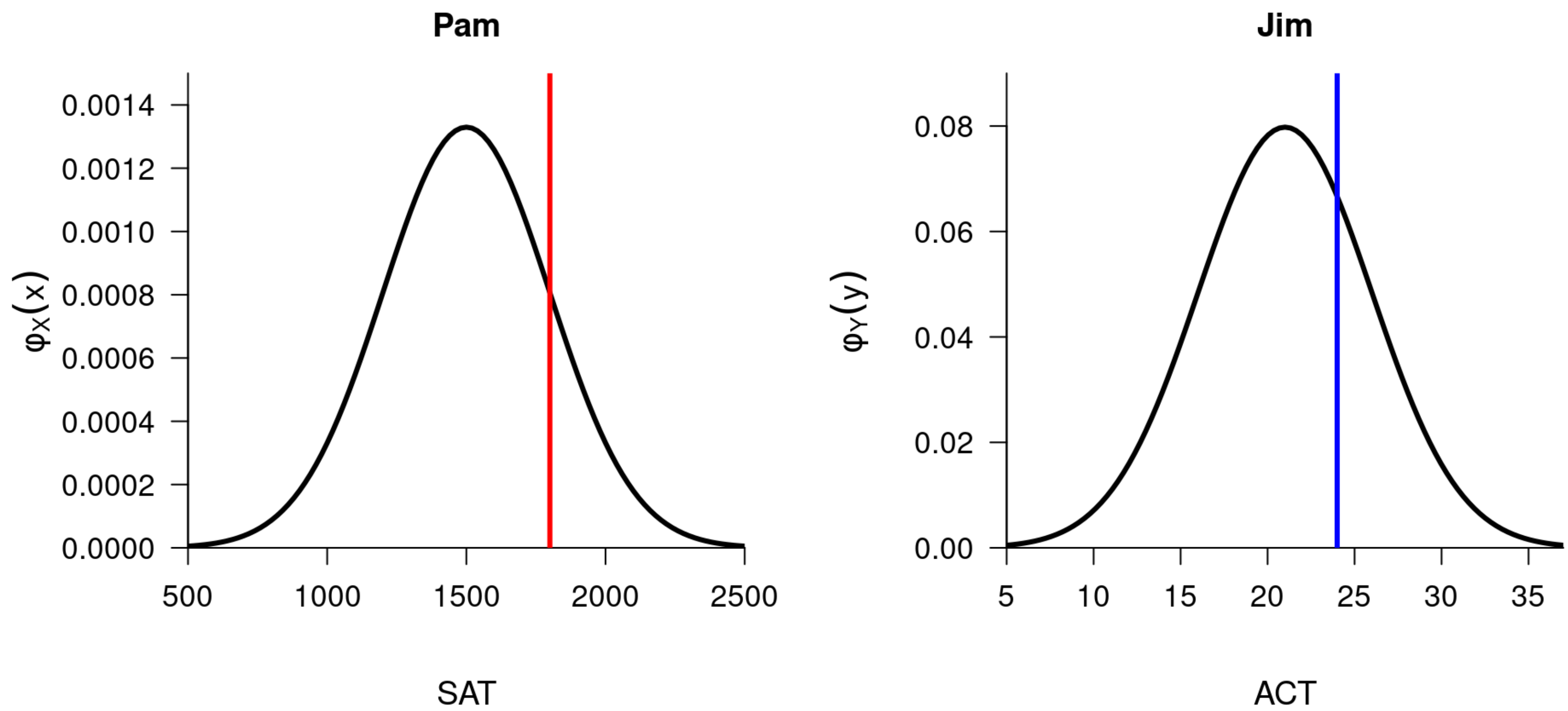
Precisamos avaliar quão melhor (ou pior) a Pam foi em relação aos demais alunos que realizaram o SAT. Precisamos avaliar quão melhor (ou pior) o Jim foi em relação aos demais alunos que realizaram o ACT.

- A universidade tem acesso à média (1500) e ao desvio-padrão (300) das notas de todos os alunos que realizaram o SAT juntamente com a Pam.
- A universidade tem acesso à média (21) e ao desvio-padrão (5) das notas de todos os alunos que realizaram o ACT juntamente com a Jim.

Assumindo que as notas dos dois testes seguem uma distribuição normal:

- Seja X uma v.a. representando a nota no SAT: $X \sim N(\mu = 1500, \sigma^2 = 300^2)$.
- Seja Y uma v.a. representando a nota no ACT: $Y \sim N(\mu = 21, \sigma^2 = 5^2)$.

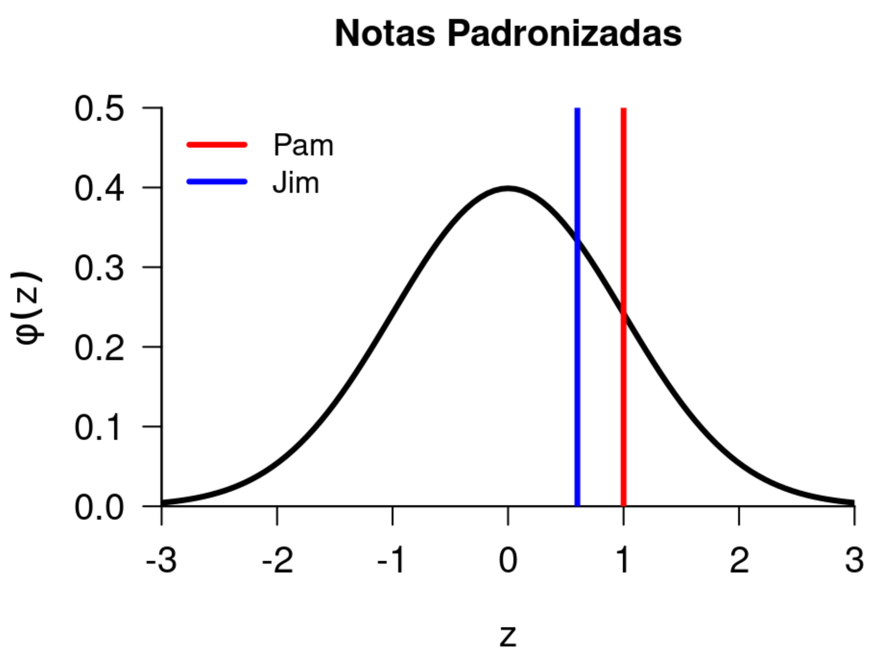
Exemplo: SAT e ACT



Seja X uma v.a. representando a nota no SAT: $X \sim N(\mu = 1500, \sigma^2 = 300^2)$. Padronizando a v.a. das notas do SAT: $Z_1 = \frac{X-1500}{300} \sim N(0,1)$. Padronizando a nota da Pam: $\frac{1800-1500}{300} = 1$.

Seja Y uma v.a. representando a nota no ACT: $Y \sim N(\mu = 21, \sigma^2 = 5^2)$. Padronizando a v.a. das notas do ACT: $Z_2 = \frac{Y-21}{5} \sim N(0,1)$. Padronizando a nota do Jim: $\frac{24-21}{5} = 0.6$.

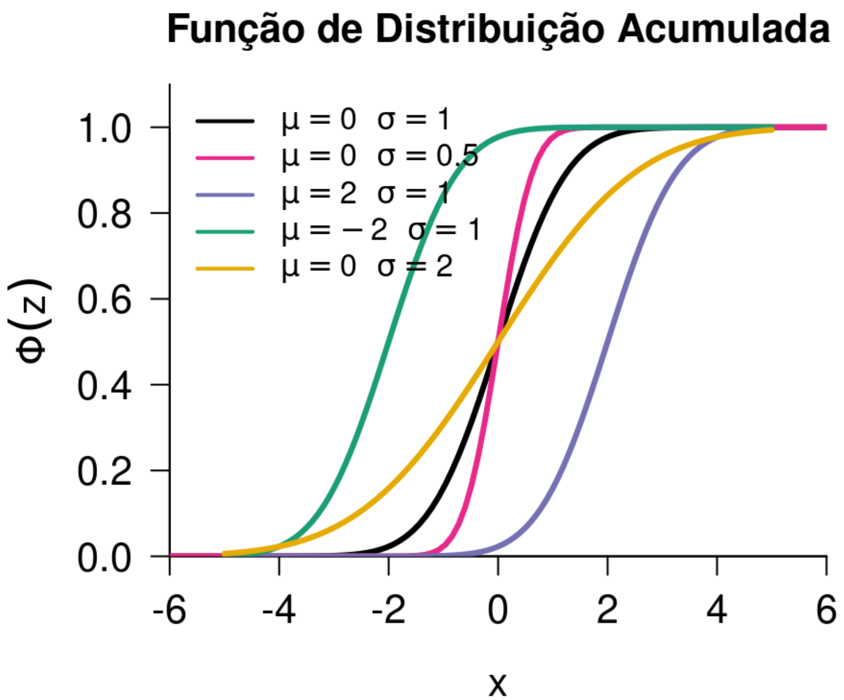
Com as notas padrozinadas, podemos compará-las:



Para calcular as probabilidades, precisamos usar a f.d.a. de $Z \sim N(0,1)$

$$\Phi(z) = P(Z \leq z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt,$$

que não tem forma fechada, pois e^{-t^2} não tem antiderivada.

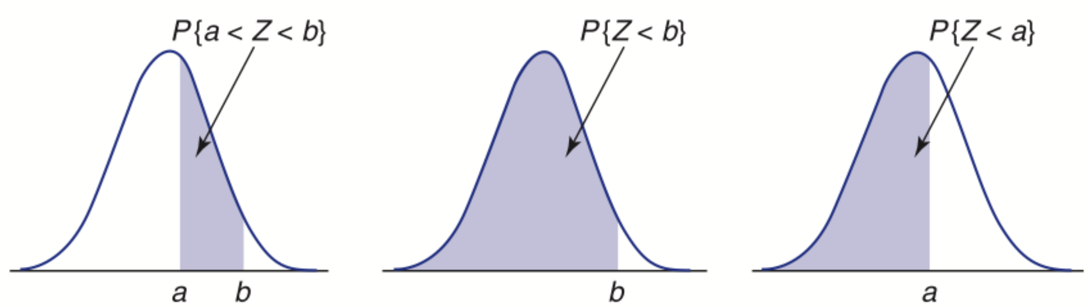


Os valores para $Z \sim N(0,1)$ e $\phi(z)$ encontram-se tabelados e/ou implementados em software como o **R**. Tudo o que precisamos fazer é transformar a variável em $N(0,1)$ e usar os valores tabelados. Ou seja, para $X \sim N(\mu, \sigma^2)$, temos:

$$F_X(a) = P(X \leq a) = P\left(\underbrace{\frac{X-\mu}{\sigma}}_{Z} \leq \frac{a-\mu}{\sigma}\right) = \Phi\left(\frac{a-\mu}{\sigma}\right).$$

A probabilidade de um intervalo é dada por:

$$\begin{aligned} P(a < Z < b) &= P(Z < b) - P(Z < a) \\ &= P(Z \leq b) - P(Z \leq a) \\ &= \Phi(b) - \Phi(a). \end{aligned}$$



Distribuição Normal

Exercitando com a tabela da Normal:

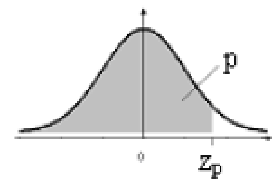
$$\Phi(0.2) = 0.5793$$

$$\Phi(0.45) = 0.6736$$

$$\Phi(1.28) = 0.8997$$

$$\Phi(-0.45) = 1 - \Phi(0.45) \\ = 0.3264$$

Tabela I: Distribuição Normal Padrão Acumulada



Fornece Φ(z) = P(-∞ < Z ≤ z), para todo z, de 0,01 em 0,01, desde z = 0,00 até z = 3,59
A distribuição de Z é Normal(0;1)

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177

O modelo de distribuição normal pode ser utilizado em experimentos nos quais seus valorem tendem a se concentrar em torno da média.

Exemplo (Magalhães e Lima, 2002). Doentes, sofrendo de certa moléstia, são submetidos a um certo tratamento intensivo cujo tempo de cura foi modelado por uma densidade Normal, de média 15 e desvio padrão 2 (em dias). Seja X o tempo de cura e, portanto,

$$X \sim N(15, 4).$$

Para sabermos a proporção de pacientes que demoram mais de 17 dias para se recuperar, temos:

$$P(X > 17) = P\left(\frac{X - 15}{\sqrt{4}} > \frac{17 - 15}{\sqrt{4}}\right) = P(Z > 1) = 0,1587$$

cujo cálculo foi feito usando a tabela da distribuição normal padrão.
A verificação da probabilidade de um paciente, escolhido ao acaso, apresentar um tempo de cura inferior a 20 dias é:

$$P(X < 20) = P\left(\frac{X - 15}{\sqrt{4}} < \frac{20 - 15}{\sqrt{4}}\right) = P(Z < 0,25) = 0,9938.$$

Procedendo de forma análoga obteremos outras probabilidades desejadas.

Outras propriedades Sejam a e b constantes conhecidas.

1. Pode ser demostrado que se X segue uma distribuição normal $X \sim \mathcal{N}(\mu, \sigma^2)$, então $aX + b \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$.
2. Se X e Y são variáveis aleatórias independentes que seguem distribuição normal, então a soma $U = X + Y$, a diferença $V = X - Y$ ou qualquer combinação linear $W = aX + bY$ também são variáveis aleatórias com distribuição normal. Em particular, se $X \sim \mathcal{N}(\mu_x, \sigma_x^2)$ e $Y \sim \mathcal{N}(\mu_y, \sigma_y^2)$ são variáveis aleatórias normais independentes então: $U = X + Y \sim \mathcal{N}(\mu_x + \mu_y, \sigma_x^2 + \sigma_y^2)$.
3. **Aproximação Normal para uma Binomial.** Seja $X \sim Bin(n, p)$. Se n é suficientemente grande, a distribuição de X pode ser aproximada pela distribuição normal, isto é,

$$X \sim N(np, np(1-p)).$$

4. **Teorema Central do Limite.** Seja $X_1, X_2 \ldots X_n$ uma sequência de variáveis aleatórias independentes com média μ e desvio padrão σ, para n suficientemente grande temos:

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

assim, no caso de n ser grande, temos que a média amostral tem distribuição normal de média μ e desvio padrão $\frac{\sigma}{\sqrt{n}}$.