Notebook

December 10, 2024

[124]:
```python
import numpy as np
import polars as pl
import seaborn as sns
import matplotlib.pyplot as plt
```

Simulate random variable X with parameter p: - $P(X = k) = p(1-p)^{(k-1)}; k \geq 1$

[ ]:
```python
def run_bernoulli(n_success: int, prob_success: float) -> int:
    """Runs a sequence of Bernoulli trials

    Args:
        n_success (int): expected number of successes before stopping trials
        prob_success (float): probability of a success in a single
            Bernoulli trial

    Returns:
        int: number of trials needed to get the requested number of sucesses
    """
    count_success = 0
    count_trials = 0
    while count_success < n_success:
        bernoulli_trial = np.random.random()
        result = bernoulli_trial < prob_success
        count_success += int(result)
        count_trials += 1

    return count_trials
```

[ ]:
```python
def run_experiment(
    n_success: int, prob_success: float, n_trials: int, n_repeat: int
) -> pl.DataFrame:
    """Runs an experiment consisting of multiple repetitions of Bernoulli Trials
    to determine a statistically relevant sample of histograms representing the
    distribution of the simulated random variable

    Args:
        n_success (int): expected number of successes before stopping trials
        prob_success (float): probability of a success in a single
```

```
            Bernoulli trial
        n_trials (int): number of Bernoulli trials to execute in a sequence
        n_repeat (int): number of repetitions of a series of trials

    Returns:
        pl.DataFrame: history of all executions in a dataframe with the columns
            - trials (int) number of trials to get a success
            - count (int) number of times the success was achievd in this
                number of trials
            - perc (float) percentage of times the success was achievd in this
                number of trials
            - repeat (int) id of the repetition
    """
    histograms = list()
    for i in range(n_repeat):
        trials_results = [
            run_bernoulli(n_success=n_success, prob_success=prob_success)
            for _ in range(n_trials)
        ]

        histograms.append(
            pl.Series(values=trials_results, name='trials')
            .value_counts()
            .sort('trials')
            .with_columns(
                perc=pl.col('count') / pl.sum('count'),
                repeat=i,
            )
        )

    return pl.concat(histograms)
```

[ ]:

```
prob_success = 0.5

results_simulated = run_experiment(
    n_success=1, prob_success=prob_success,
    n_trials=1000, n_repeat=20,
)

trials_max_sim = results_simulated.select(pl.max('trials')).item()
results_theoretic = [
    prob_success * (1 - prob_success) ** (k - 1)
    for k in range(1, trials_max_sim)
]
```

[193]:

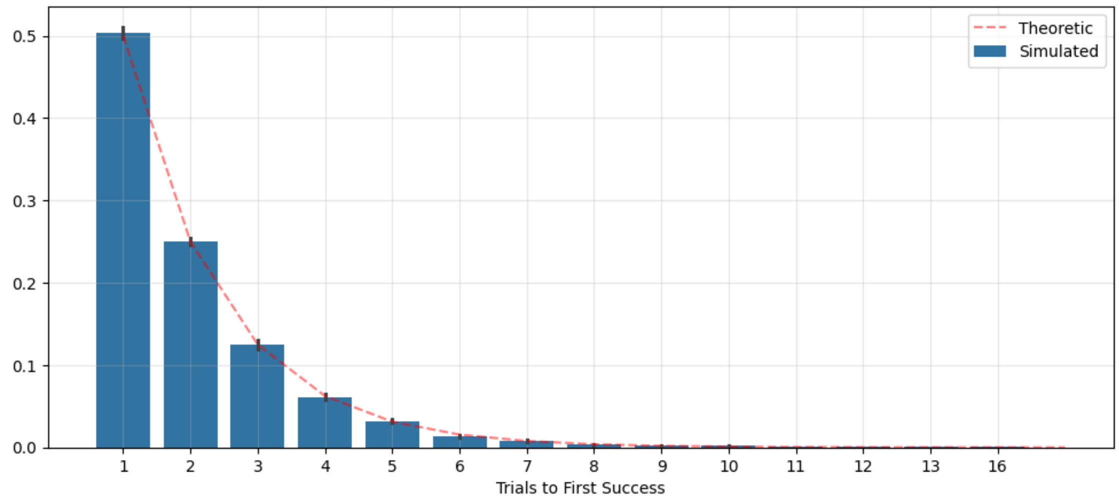
```
plt.figure(figsize=(12, 5))
ax = plt.gca()
```

```
sns.barplot(
    results_simulated,
    x='trials', y='perc',
    errorbar=('ci', 95), n_boot=1000,
    label='Simulated',
    ax=ax,
)

sns.lineplot(
    results_theoretic,
    color='red', ls='--', alpha=0.5,
    label='Theoretic',
    ax=ax,
)

ax.set_xlabel('Trials to First Success')
ax.set_ylabel('')
ax.grid(alpha=0.3)

plt.show()
plt.close()
```



This notebook was converted with convert.ploomber.io

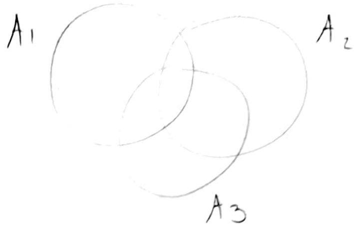
1) $P\left(\bigcup_{i=1}^{n} A_1\right) \leq \sum_{i=1}^{n} P(A_i)$

$* \; P(A \cup B) = P(A) + P(B) - P(A \cap B)$

$i = 1 \to P(A_1) \leq P(A_1)$

$i = 2 \to \underbrace{P(A_1 \cup A_2)} \leq P(A_1) + P(A_2)$

$$P(A_1) + P(A_2) - \underbrace{P(A_1 \cap A_2)}_{\geq 0} \to P(A_1 \cup A_2) \leq P(A_1) + P(A_2)$$



$i = 3 \to P(A_1 \cup A_2 \cup A_3)$

$$= P(A_1) + P(A_2) + P(A_3) \underbrace{- P(A_1 \cap A_2) - P(A_2 \cap A_3) - P(A_1 \cap A_3)}_{\geq - P(A_1 \cap A_2)} + P(A_1 \cap A_2 \cap A_3)$$

dado $P(A_2 \cap A_3) \geq 0$

$P(A_1 \cap A_3) \geq 0$

$\downarrow$

$\geq - P(A_1 \cap A_2 \cap A_3) \to P(A_1 \cup A_2 \cup A_3) \leq P(A_1) + P(A_2) + P(A_3)$

$* \; (A \cup B \cup C) = (A \cup B) \cup C$

$* \; (A \cap B \cap C) = (A \cap B) \cap C$

$$P\left(\bigcup_{i=1}^{n} A_i\right) = P\left(\bigcup_{i=1}^{n-1} A_i\right) + P(A_n) - \underbrace{P\left(\bigcup_{i=1}^{n-1} A_i \cap A_n\right)}_{\geq 0}$$

$$P\left(\bigcup_{i=1}^{n} A_i\right) \leq \underbrace{P\left(\bigcup_{i=1}^{n-1} A_i\right)}_{\leq P\left(\bigcup_{i=1}^{n-2} A_i\right) + P(A_{n-1})} + P(A_n)$$

portanto

$$\boxed{P\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} P(A_i)}$$

2) $\Omega = \{A, B, C\}$

$P(A) = p$; $P(B) = q$; $P(C) = r$

* $p + b + c = 1$

$P(A)$ antes de B ocorrer $\Rightarrow P(A \mid \bar{B})$

* $P(x \mid y) = P(x \cap y) / P(y)$

$P(A \mid \bar{B}) = P(A \cap \bar{B}) / P(\bar{B})$

↳ $P(A \cap \bar{B}) = P(A)$

↳ $P(\bar{B}) = P(A) + P(C)$

$\rightarrow \boxed{P(A \mid \bar{B}) = p/p+c}$

3) $P(X \geq x+t \mid X \geq t) = P(X \geq x) \quad x, t > 0$

$P(X \geq x+t \mid X \geq t) = P(X \geq x+t \cap X \geq t) / P(X \geq t)$

$* P(X \geq x+t) \subset P(X \geq t)$

$\rightarrow = P(X \geq x+t)$

$\rightarrow = P(X \geq x+t) / P(X \geq t)$

$*$ CDF $\rightarrow F(X = x) = P(X < x)$

$1 - F(X = x) = P(X \geq x)$

$$\left.\begin{aligned} P(X \geq t) &= 1 - F(X = t) = G(t) \\ P(X \geq x+t) &= 1 - F(X = x+t) = G(x+t) \\ P(X \geq x) &= 1 - F(X = x) = G(x) \end{aligned}\right\} \quad G(x) = \frac{G(x+t)}{G(t)}$$

$* \quad G(x+t) = G(x) \cdot G(t) \rightarrow G = ?$

$\rightarrow G(a+a) = G(a) \cdot G(a) = G(a)^2$

$G(3a) = G(a)^3 \longrightarrow G(ca) = G(a)^c$

$G\left(\frac{1}{c} a\right) = G(a)^{1/c}$

$G\left(\frac{c_1}{c_2} a\right) = G(a)^{c_1/c_2}$

$* \quad G(xa) = G(a)^x$

$a = 1$

$\rightarrow G(x) = G(1)^x \rightarrow \left[e^{\ln(G(1))}\right]^x = G(1)^x = e^{\underbrace{x \ln G(1)}}$

$G(x) \geq 0 \ \forall x$

$\ln G(1) = -\lambda$

$G(x) = e^{-\lambda x} \quad (\mathrm{I}) ; \quad x > 0 ; \lambda > 0$

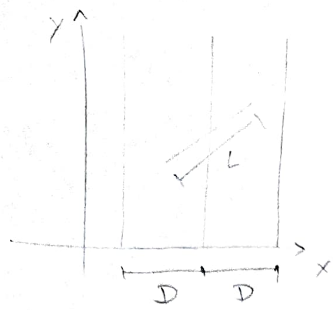
$Y \sim \exp \rightarrow$ CDF $\Rightarrow 1 - e^{-\lambda x}$

$1 - \text{CDF} = G(x) = 1 - (1 - e^{-\lambda x}) = e^{-\lambda x} \quad (\mathrm{II})$

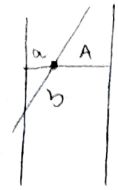
$(\mathrm{I}) = (\mathrm{II})$

CDF de uma distr. exp. é igual a CDF de uma distr. sem memória

4)



$L \geq D$

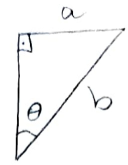


$\boxed{a + A = D}$

* distância do centro da agulha

$a \to$ até a linha mais próxima

$A \to$ até a linha mais distante



$\boxed{\operatorname{sen}\theta = \frac{a}{b}} \to \boxed{b = a/\operatorname{sen}\theta}$

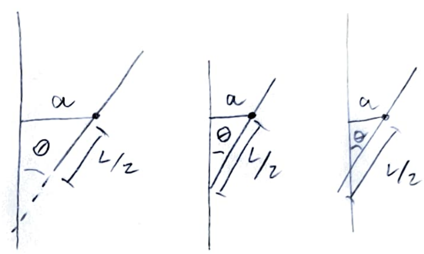
* no caso que o centro da agulha está no meio entre as linhas

$a = A \to \boxed{a \leq D/2}$

θ	a	cruza
$=0$	$=0$	sim
$=0$	>0	não
>0	$=0$	sim
>0	>0	?

? $\to a/\operatorname{sen}\theta \leq L/2$ sim

$\to a/\operatorname{sen}\theta > L/2$ não



$$P(a) \sim \text{Unif}(0, D/2) \to \int_0^{D/2} c\,da = 1 \to \boxed{c_a = 2/D}$$

$$P(\theta) \sim \text{Unif}(0, \pi/2) \to \int_0^{\pi/2} c\,d\theta = 1 \to \boxed{c_\theta = 2/\pi}$$

supondo independência:

$$P(a,\theta) = P(a) \cdot P(\theta) = 4/D\pi \; ; \quad 0 \leq a \leq D/2 \; ; \quad 0 \leq \theta \leq \pi/2$$

$$P(\text{cruza}) = \int_{\theta=0}^{\pi/2} \int_{x=0}^{\frac{L\operatorname{sen}\theta}{2}} 4/D\pi \, dx\, d\theta = \frac{4}{D\pi} \int_{\theta=0}^{\pi/2} L\operatorname{sen}\theta/2 \, d\theta$$

$$= \frac{4}{D\pi} \cdot \frac{L}{2}\left(-\cos(\pi/2) + \cos(0)\right) = \boxed{\frac{2L}{D\pi}}$$

5) $X = \cos\theta \qquad Y = \operatorname{sen}\theta$

$\theta \sim \text{unif}(0, 2\pi)$

a) se $\operatorname{cov}(A,B) = 0 \rightarrow A$ e B são não correlacionadas

* $\operatorname{cov}(A,B) = E_{AB}(AB) - E_A(A)\,E_B(B)$

* $B = g(A) \rightarrow E(g(a)) = \int_{-\infty}^{\infty} g(a)\,P_A(a)\,da$

$$P_\theta(\theta) = \begin{cases} 1 & 0 < \theta < 2\pi \\ 0 & cc \end{cases}$$

$$E(\cos(\theta)) = \int_0^{2\pi} \cos(\theta)\cdot 1\,d\theta = \cancel{\operatorname{sen}(2\pi)}^{\,0} - \cancel{\operatorname{sen}(0)}^{\,0} = 0$$

$$E(\operatorname{sen}(\theta)) = \int_0^{2\pi} \operatorname{sen}(\theta)\cdot 1\,d\theta = -\cancel{\cos(2\pi)}^{\,1} - (-\cancel{\cos(0)}^{\,1}) = 0$$

$$\boxed{E_x(x) = 0 \;;\; E_y(y) = 0}$$

$$E(XY) = \int_0^{2\pi} \cos\theta \operatorname{sen}\theta \cdot 1\,d\theta = -\frac{\cos(x)^2}{2}\Big|_0^{2\pi} = -1-(-1) = \boxed{0}$$

$$\boxed{\operatorname{cov}(XY) = 0}$$

b) se $P_{X,Y}(X,Y) = P_X(X)\cdot P_Y(Y)$ então são indep

8) a) $X \sim Unif(0,1)$ $\quad Y \sim Unif(0,1)$ $\quad E(U) = ?$

$U = \min(X,Y)$ $\quad V = \max(X,Y)$ $\quad cov(U,V) = ?$

$P_U(U \le u) = P_U(\min(X,Y) \le u) = 1 - P_U(\underbrace{\min(X,Y) > u}_{u<X\ ;\ u<Y})$

$= 1 - P_U(u < X, u < Y)$

$= 1 - P_U(u < X) \cdot P_U(u < Y)$

* $A \sim unif(0,1) \to P_A(a < c) = 1 - P_A(c \le a) = 1 - a$

$P_U(U \le u) = 1 - (1-u)(1-u) = \boxed{1 - (1-u)^2}$

* $E(x) = \int_{-\infty}^{\infty} x P_x(x)\,dx$

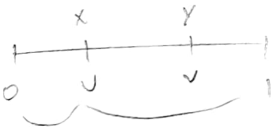
$F(u) = 1 - (1-u)^2 \to \boxed{P(u) = 2(1-u)}$

$E(U) = \int_0^1 u[2(1-u)]\,du = 2\int_0^1 (u - u^2)\,du = 2\left[\frac{x^2}{2} - \frac{x^3}{3}\right]\Big|_0^1 = \boxed{1/3}$

* $cov(X,Y) = E_{XY}(XY) - E_X(X)E_Y(Y)$

$cov(U,V) = E_{UV}(UV) - E_U(U)E_V(V)$

$X + Y = \min(X,Y) + \max(X,Y)$



$E(X) = E(Y) = 1/2$

$E(X) + E(Y) = E(U) + E(V)$

$E(U) = 1/3 \to 1/2 + 1/2 = 1/3 + E(V) \to \boxed{E(V) = 2/3}$

$U \cdot V = X \cdot Y$

$E(UV) = E(XY)$

$= E(X)E(Y) = 1/2 \cdot 1/2 = \boxed{1/4}$

$cov(U,V) = 1/4 - 1/3 \cdot 2/3 = \boxed{1/36}$

b) $P_{XY}(x,y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \cdot \exp\left(-\frac{1}{2(1-\rho^2)} \cdot (x^2 - 2\rho xy + y^2)\right)$

$z = \frac{(Y - \rho X)}{\sqrt{1-\rho^2}}$ $\quad X \perp\!\!\!\perp Z$ $\quad X \sim \text{Norm}(0,1)$, $Z \sim \text{Norm}(0,1)$ ✓ mostre que

deduza

$P(X>0, Y>0) = \frac{1}{4} + \frac{1}{2\pi} \cdot \arcsen(\rho)$

$P_X(x) = \int_{-\infty}^{\infty} P_{XY}(x,y)\,dy$

$P_X(x) = \int_{-\infty}^{\infty} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[\frac{-1}{2(1-\rho^2)} \underbrace{(x^2 - 2\rho xy + y^2)}\right]$

$+\rho^2x^2 - \rho^2x^2 \Rightarrow (y-\rho x)^2 + (1-\rho^2)x^2$

$P_X(x) = \int_{-\infty}^{\infty} \underbrace{\frac{1}{2\pi\sqrt{1-\rho^2}}}_{\sqrt{2\pi}\sqrt{2\pi(1-\rho^2)}} \exp\left[\frac{-1}{2(1-\rho^2)} \left((y-\rho x)^2 + \underbrace{(1-\rho^2)x^2}_{\text{indep. } y}\right)\right] dy$

$= \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \int_{-\infty}^{\infty} \frac{1}{\underbrace{\sqrt{2\pi(1-\rho^2)}}_{\sigma^2}} \exp\left[\frac{-1}{2\underbrace{(1-\rho^2)}_{\sigma^2}} (y - \underbrace{\rho x}_{\mu})^2\right] dy$

$\Downarrow$ PDF ~ Norm

$* \int_{-\infty}^{\infty} P_X(x)\,dx = 1$

$P_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} = \frac{1}{1\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-0}{1}\right)^2} \to \text{Norm}(0,1)$

$\hookrightarrow X \sim \text{Norm}(0,1)$

$Y \sim \text{Norm}(0,1)$

$\to E(z) = E\left[\frac{(Y-\rho X)}{\sqrt{1-\rho^2}}\right] = \frac{E(\overset{0}{Y}) - \rho E(\overset{0}{X})}{\sqrt{1-\rho^2}}$

$\boxed{= 0}$

$\to \text{Var}(z) = E(z^2) - E(z)^2$ (with $E(z)^2 = 0$)

$E(z^2) = E\left[\left(\frac{Y-\rho X}{\sqrt{1-\rho^2}}\right)^2\right] = \frac{E[\overbrace{(Y-\rho X)^2}^{Y^2 - 2\rho XY + \rho^2 X^2}]}{1-\rho^2} = \frac{E(\overset{1}{Y^2}) - 2\rho E(\overset{\rho}{XY}) + \rho^2 E(\overset{1}{X^2})}{1-\rho^2}$

$\rho = \text{cov}(X,Y)/\cancel{\sigma_X}\cancel{\sigma_Y} = E(XY) - E(\overset{0}{X})E(\overset{0}{Y}) \to \rho = E(XY)$

$E(z^2) = (1 - 2\rho^2 + \rho^2)/(1-\rho^2) = \boxed{1} \Rightarrow \boxed{Z \sim \text{Norm}(0,1)}$

* se $cov(XZ) = 0$ então são indep

$$cov(XZ) = E(XZ) - \cancel{E(X)}^{0} \cancel{E(Z)}^{0}$$

$$E(XZ) = E\left[X \frac{(Y - \rho X)}{\sqrt{1-\rho^2}}\right] = \frac{\cancel{E(XY)}^{\rho} - \rho \cancel{E(X^2)}^{1}}{\sqrt{1-\rho^2}} = \frac{\rho - \rho}{\sqrt{1-\rho^2}} = \boxed{0}$$

9) pessoa $\rightarrow$ 2 genes $\rightarrow$ N || C

NN || NC || CN $\rightarrow$ normal

CC $\rightarrow$ fibrose

$$\frac{G_1 G_2}{P_1} + \frac{G_3 G_4}{P_2} \;\Big\downarrow\; \frac{G_x G_y}{C_x}$$

a) $P_1 \rightarrow \{NN, NC, CN\}$

$P_2 \nearrow$

$C_1 \rightarrow \{NN, NC, CN\}$

$C_2 \rightarrow \{CC\} \rightarrow P_1$ e $P_2 \rightarrow \{NC, CN\}$

$NC \times NC = NN \,||\, NC \,||\, CN \,||\, CC$

$$\boxed{P(C_1 \in \{NC, CN\}) = 2/3}$$

b) $\frac{n(N)}{n(C)} = \frac{49}{1}$ $\quad * \; P(cx \mid \overline{cc}) = P(cx \cap \overline{cc}) / P(\overline{cc})$

$\hookrightarrow P(N) = 49/50$

$P(C) = 1/50$

dada independência:

$P(NN) = P(N) \cdot P(N) = 49/50 \cdot 49/50$

$P(NC) = P(N) \cdot P(C) = 49/50 \cdot 1/50$

$P(CC) = P(C) \cdot P(C) = 1/50 \cdot 1/50$

$P(cx \cap \overline{cc}) = P(N) \cdot P(C) + P(C) \cdot P(N)$

$= 2 \cdot 49/50 \cdot 1/50 = \boxed{0{,}0392}$

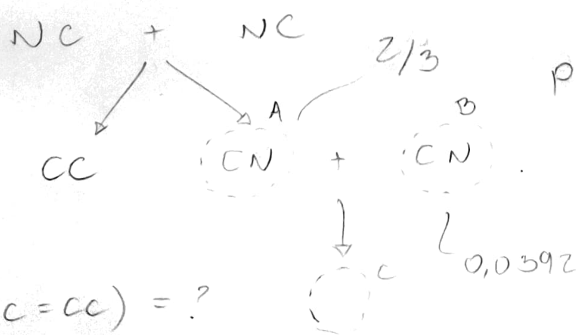
$P(\overline{cc}) = 1 - P(C) \cdot P(C)$

$= 1 - 1/50 \cdot 1/50 = \boxed{0{,}9996}$

$P(cx \mid \overline{cc}) = \frac{0{,}0392}{0{,}9996}$

$\boxed{\approx 0{,}0392}$

c)



$P(c = cc) = ?$

$CN \times CN \Rightarrow CC \;\; CN \;\; NC \;\; CC$

$P(c = cc) = P(A = cx)$

$\cdot \, P(B = cx)$

$\cdot \, P(c = cc \mid A = cx, B = cx)$

$= 2/3 \cdot 0{,}0392 \cdot 1/4$

$\boxed{= 0{,}0065}$