Fundação Getúlio Vargas

Contributors: Jean² e Gabriel Matos
Content: Teoria da Probabilidade

Este material contempla um resumo dos principais tópicos da Teoria de Probabilidade, abordando as definições e os resultados necessários.

Para quaisquer correções ou apontamentos acerca desse material, não exitem em nos contatar, seguem abaixo os nossos emails:

- jeannrochasilva@gmail.com
- gplaygmg775@gmail.com

Contagem

Seguem abaixo algumas Ferramentas Úteis e Essenciais para a Contagem...

1. Princípio Aditivo
2. Princípio Fundamental da Contagem (ou Princípio Multiplicativo)
3. Permutações Simples
4. Arranjos
5. Permutações com Repetição
6. Permutações Circulares
7. Combinações Simples
8. Combinações Completas (ou Combinações com Repetiação, ou ainda Modelo de Pauzinhos e Bolinhas)
9. Permutações Caóticas
10. Contagem Dupla (ou Story Proofs)
11. Princípio da Indução Finita

Probabilidade

Definição 1 (Probabilidade: Caso Equiprorável)**.** Seja $A \subset S$, com S finito. Então, a probabilidade de A é

$$P(A) = \frac{|A|}{|S|}$$

Comentário: Esta definição só faz sentido em espaços finitos equiproráveis.

Definição 2 (Probabilidade: Geral)**.** Sendo $\mathcal{P}(S)$ o conjunto de todos os subconjuntos de S, uma probabilidade em S é uma função $P : \mathcal{P}(S) \to [0, 1]$ tal que

1. $P(\emptyset) = 0$
2. $P(S) = 1$
3. Se $A_1, A_2, ... \subset S$ são disjuntos, então

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n)$$

Proposição 1. Seja P uma probabilidade em S. Então

1. $P(A^C) = 1 - P(A), \forall A \subset S$
2. $A \subset B \Rightarrow P(A) \leq P(B), \forall A, B \subset S$
3. $P(A \cup B) = P(A) + P(B) - P(A \cap B), \forall A, B \subset S$

A útlima propriedade pode ser generalizada para

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i} P(A_i) - \sum_{i<j} P(A_i \cap A_j) + \sum_{i<j<k} P(A_i \cap A_j \cap A_k) - ... + (-1)^{n+1} P(A_1 \cap ... \cap A_n)$$

Probabilidade Condicional

Definição 3. Dados $A, B \subset S$ tal que $P(B) > 0$, a probabilidade condicional de A dado que B ocorreu (ou, simplesmente, a probabilidade de B dado A) é

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Proposição 2. Dado $B \subset S$, a função $P(\cdot|B)$ é uma probabilidade.

Proposição 3 (Teorema de Bayes)**.** Sejam $A_1, ..., A_n$ uma partição de S, Então

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B)} = \frac{P(B|A_i)P(A_i)}{\sum\limits_{i=1}^{n} P(B|A_j)P(A_j)}, \forall B \subset S$$

Em particular, como A, A^C forma uma partição de S, temos

$$P(A_i|B) = \frac{P(B|A)P(A)}{P(B)} = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^C)P(A^C)}, \forall B \subset S$$

Proposição 4 (Teorema da Probabilidade Total)**.** Sejam $A_1, ..., A_n$ uma partição de S. Então

$$P(B) = P(B|A_1)P(A_1) + ... + P(B|A_n)P(A_n), \forall B \subset S$$

Proposição 5. O Teorema de Bayes e o Teorema da Probabilidade Total admitem as seguintes versões quando o condicionamento é dado por uma intersecção de eventos

1. **Teorema de Bayes:**

$$P(A_i|B, C) = \frac{P(B|A_i, C)P(A_i|C)}{P(B|C)} = \frac{P(B|A_i, C)P(A_i|C)}{\sum\limits_{j=1}^{n} P(B|A_j, C)P(A_j|C)}, \forall B \subset S$$

2. **Teorema da Probabilidade Total**

$$P(B|C) = P(B|A_1, C)P(A_1|C) + ... + P(B|A_n, C)P(A_n|C), \forall B \subset S$$

Variáveis Aleatórias Discretas

Definição 4. Seja $S = \{s_1, s_2, ..., s_n\}$ (ou $S = \{s_1, s_2, ...\}$) um espaço amostral. Uma variável aleatória é uma função $X : S \to \mathbb{R}$.

Definição 5. Uma variável aleatória X é dita discreta se a sua imagem é da forma $X(S) = \{a_1, a_2, ..., a_n\}$ (ou $X(S) = \{a_1, a_2, ...\}$). O conjunto $X(S)$ é dito suporte e denotado $\text{Supp}(X)$.

Distribuições

- **Função de Massa de Probabilidade (Probability Mass Function)**

  Definição 6. A função de massa probabilidade de uma variável aleatória discreta X é a função $p_X : \mathbb{R} \to \mathbb{R}$ dada por $p_X(x) = P(X = x)$.
  Comentário: Para espaços equiproráveis, isso se assemelha a uma frequência relativa, pois estamos preocupados em analisar a probabilidade que um dado valor x assume no espaço de possibilidades.

Proposição 6. Seja X uma variável aleatória discreta. Então

(a) $p_X(x) > 0, \forall x \in \text{Supp}(X)$ e $p_X(x) = 0, \forall x \notin \text{Supp}(X)$

(b) $\sum\limits_{x \in \text{Supp}(X)} p_X = 1$

- **Função de Probalidade Acumulada (Acumulative Probability Function)**

  Definição 7. A função de probabilidade acumulada de uma variável aleatória discreta X é a função $F_X : \mathbb{R} \to \mathbb{R}$ dada por $F_X(x) = P(X \leq x)$.

  Proposição 7. Seja X uma variável aleatória. Então

  (a) $x_1 \leq x_2 \Rightarrow F_X(x_1) \leq F_X(x_2)$ (não decrescente)

  (b) $\lim\limits_{x \to -\infty} F_X(x) = 0$ e $\lim\limits_{x \to +\infty} F_X(x) = 1$

  (c) $\lim_{x \to a^+} F_X(x) = F(a)$ (contínua à direita)

Exemplos de Distribuições Discretas

- **Bernoulli:** Experimento, onde a probabilidade de sucesso é p e a de fracasso é $1 - p$.

$$X = \begin{cases} 0 & \text{se obteve fracasso} \\ 1 & \text{se obteve sucesso} \end{cases}, \qquad p_X(k) = \begin{cases} 1 - p & \text{se } k = 0 \\ p & \text{se } k = 1 \end{cases}$$

  Notação: $X \sim \text{Bern}(p)$

- **Binomial:** n eventos independentes de Bernoulli com mesma probabilidade.

$$X = \text{nº de sucessos}, \qquad p_X(k) = \tbinom{n}{k} p^k (1-p)^{n-k}$$

  Notação: $X \sim \text{Bin}(n, p)$

  Proposição 8. $X \sim \text{Bin}(n, p) \Rightarrow n - X \sim \text{Bin}(n, 1 - p)$

- **Hipergeométrica:** n eventos (dependentes) de Bernoulli em uma amostra, na qual o i-ésimo evento representa uma retirada da amostra (sem reposição).

$$X = \text{nº de sucessos}, \qquad p_X(k) = \frac{\binom{n}{k}\binom{m}{t-k}}{\binom{m+n}{t}}$$

  Notação: $X \sim \text{HGeom}(m, n, t)$
  Comentário: Caso houvesse reposição, teríamos n eventos independentes de Bernoulli e, portanto, uma binomial.

Proposição 9. $X \sim \text{HGeom}(m, n, t) \sim \text{HGeom}(t, m+n-t, m)$

- **Uniforme:** espaço equiprorável (S)

$$X = \text{n}^{\underline{\text{o}}} \text{ de sucessos em } C \subset S, \qquad p_X(x) = \frac{1}{|C|}$$

Notação: $X \sim \text{DUnif}(C)$

- **Geométrica:** Eventos independentes de Bernoulli com mesma probabilidade.

$$X = \text{n}^{\underline{\text{o}}} \text{ de falhas até o } 1^{\underline{\text{o}}} \text{ sucesso}, \qquad p_X(k) = (1-p)^k p$$

Notação: $\text{Geom}(p)$

- **Primeiro Sucesso:** Eventos independentes de Bernoulli com mesma probabilidade.

$$X = \text{n}^{\underline{\text{o}}} \text{ de tentativas até o } 1^{\underline{\text{o}}} \text{ sucesso, incluindo este}, \qquad p_X(k) = (1-p)^{k-1} p$$

Notação: $\text{FS}(p)$

- **Binomial Negativa:** Eventos independentes de bernoulli com mesma probabilidade.

$$X = \text{n}^{\underline{\text{o}}} \text{ de falhas até o } k^{\underline{\text{o}}}\text{-ésimo sucesso}, \qquad p_X(k) = \tbinom{k+n-1}{n-1}(1-p)^n p^k$$

Notação: $\text{NBin}(n, p)$

- **Poisson:** Eventos que ocorrem de forma independente a uma taxa média λ.

$$X_t = \text{n}^{\underline{\text{o}}} \text{ de ocorrências em } [0, t), \qquad p_{X_t}(k) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}$$

Notação: $\text{Pois}(\lambda)$

Funções de Variáveis Aleatórias

Definição 8 (Não é bem uma definição)**.** Seja $g : \mathbb{R} \to \mathbb{R}$ uma função e X uma variável aleatória. Então, a função $g(X) = g \circ X$ é a variável aleatória que mapeia $s \in S$ em $g(X(s)) \in \mathbb{R}$.

Proposição 10. Se X é uma variável aleatória discreta, então $g(X)$ é uma variável aleatória discreta, $\forall g : \mathbb{R} \to \mathbb{R}$ função.

Proposição 11. Seja X uma variável aleatória e $g : \mathbb{R} \to \mathbb{R}$ uma função. Então

$$p_{g(X)}(y) = \sum_{x:g(x)=y} P(X = x)$$

Em particular, se g é bijetiva, então

$$p_{g(X)}(y) = P(X = g^{-1}(y))$$

Definição 9 (Não é bem uma definição)**.** Seja $g : \mathbb{R}^2 \to \mathbb{R}$ uma função e X e Y variáveis aleatórias. Então, a função $g(X, Y) = g \circ (X, Y)$ é a variável aleatória que mapeia $s \in S$ em $g(X(s), Y(s)) \in \mathbb{R}$.
Comentário: É fácil generalizar essa ideia para n variáveis aleatórias.

Conexões entre as Distribuições

Definição 10. Para quaisquer variáveis aleatórias X e Z definimos a PMF condicional de X fixado $Z = z$ como $P(X = x|Z = z)$. Note que z é fixo, de maneira que a quantidade $P(X = x|Z = z)$ é tratada como uma função de x.

- **Binomial e Hipergeométrica:**
  - Se X e Y são tais que $X \sim \text{Bin}(m, p)$ e $Y \sim \text{Bin}(n, p)$, então a distribuição condicional de X dado $X + Y = t$ é $\text{HGeom}(m, n, t)$.
  - Se $X \sim \text{HGeom}(m, n, t)$ e $m + n \to +\infty$ e $p = \dfrac{m}{m + n}$ permanece fixo, então a distribuição de X converge para $\text{Bin}(t, p)$.
- **Binomial e Poisson**
  - Se X e Y são tais que $X \sim \text{Pois}(\lambda_1)$ e $Y \sim \text{Pois}(\lambda_2)$, então a distribuição condicional de X dado $X + Y = n$ é $\text{Bin}(n, \lambda_1/(\lambda_1 + \lambda_2))$.
  - Se $X \sim \text{Bin}(n, p)$ e $n \to +\infty$ e $\lambda = np$ permanece fixo, então a distribuição de X converge para $\text{Pois}(\lambda)$.

Independência

Definição 11. Dois eventos A e B são ditos independentes se

$$P(A \cap B) = P(A)P(B)$$

Se $P(A) > 0$ e $P(B) > 0$, isto é equivalente a $P(A|B) = P(A)$ e $P(B|A) = P(B)$, ou seja, tanto a probabilidade de A quanto a de B ocorrerem não sofre influência uma da outra. Mais geralmente, $A_1, A_2, ..., A_n$ (ou $A_1, A_2, ...$) eventos são independentes se para toda coleção $\{i_1, ..., i_m\}$ finita, temos

$$P(A_{i_1} \cap A_{i_2} \cap ... \cap A_{i_m}) = P(A_{i_1})...P(A_{i_m})$$

Comentário: É recomendado conhecer algum exemplo de 3 eventos A, B e C tais que $P(A \cap B \cap C) = P(A)P(B)P(C)$, mas $P(A \cap B) \neq P(A)P(B)$ ou $P(A \cap C) \neq P(A)P(C)$ ou $P(B \cap C) \neq P(B)P(C)$, e outros 3 eventos E, F e G tais que $P(E \cap F) = P(E)P(F)$, $P(E \cap G) = P(E)P(G)$, $P(F \cap G) = P(F)P(G)$, mas $P(E \cap F \cap G) \neq P(E)P(F)P(G)$.

Proposição 12. Se A e B são independentes, então A e B^C, A^C e B, e A^C e B^C também são

Definição 12 (Condicionalmente Independente)**.** Dois eventos A e B são ditos independentes condicionalmente dado E se

$$P(A \cap B|E) = P(A|E)P(B|E)$$

Comentário: Nem sempre independência implica condicionalmente independente e vice versa.

Definição 13. X e Y são variáveis aleatórias independentes se

$$P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y), \forall x, y \in \mathbb{R}$$

No caso discreto, isto é equivalente a

$$P(X = x, Y = y) = P(X = x)P(Y = y), \forall x, y \in \mathbb{R}$$

Mais geralmente, $X_1, ..., X_n$ são variáveis aleatórias independentes se

$$P(X_1 \leq x_1, ..., X_n \leq x_n) = P(X_1 \leq x_1)...P(X_n \leq x_n), \forall x, y \in \mathbb{R}$$

Proposição 13. Se $X_1, X_2, ..., X_n$ (ou $X_1, X_2, ...$) são variáveis aleatórias independentes, então qualquer coleção $X_{i_1}, ..., X_{i_m}$ finita é independente.

Definição 14. X e Y são variáveis aleatórias identicamente distribuídas se $F_X = F_Y$. No caso discreto, isto equivale a $p_X = p_Y$.
Comentário: É recomendado conhecer algum exemplo de variáveis aleatórias X e Y que são independentes e identicamente distribuídas, que são independentes, mas não identicamente distribuídas, que são identicamente distribuídas, mas não independentes e, que não são nem independentes e nem identicamente distribuídas.

Proposição 14. Sejam $X_1, ..., X_n$ distribuições de Bernoulli independentes e identicamente distíbuidas, com distribuição $\sim \text{Bern}(p)$. Então, $X = X_1 + ... + X_n \sim \text{Bin}(n, p)$.

Proposição 15. Se X e Y são independentes e tais que $X \sim \text{Bin}(n, p)$ e $Y \sim \text{Bin}(m, p)$, então $X + Y \sim \text{Bin}(m + n, p)$.

Proposição 16. Sejam $X_1, ..., X_n$ distribuições Geométricas independentes e identicamente distíbuidas, com distribuição $\sim \text{Geom}(p)$. Então, $X = X_1 + ... + X_n \sim \text{NBin}(n, p)$.

Proposição 17. Se X e Y são independentes são tais que $X \sim \text{Pois}(\lambda_1)$ e $Y \sim \text{Pois}(\lambda_2)$, então $X + Y \sim \text{Pois}(\lambda_1 + \lambda_2)$

Definição 15. X e Y são variáveis aleatórias condicionalmente independentes dado uma variável aleatória Z se

$$P(X \leq x, Y \leq y | Z = z) = P(X \leq x | Z = z) P(Y \leq y | Z = z), \forall x, y \in \mathbb{R} \text{ e } \forall z \in \text{Supp}(Z)$$

No caso discreto, tem-se a seguinte definição equivalente

$$P(X = x, Y = y | Z = z) = P(X = x | Z = z) P(Y = y | Z = z), \forall x, y \in \mathbb{R} \text{ e } \forall z \in \text{Supp}(Z)$$

Esperança

Definição 16. O valor esperado de uma variável aleatória discreta X é

$$E(X) = \sum_{x \in \text{Supp}(X)} x P(X = x)$$

Proposição 18. Se X e Y são discretas com a mesma distribuição, então $E(X) = E(Y)$.

Proposição 19. Sejam X e Y variáveis aleatórias e $\lambda \in \mathbb{R}$. Então

$$E(X + \lambda Y) = E(X) + \lambda E(Y)$$

Proposição 20. Se X e Y são variáveis aleatórias tais que $P(X \geq Y) = 1$, então $E(X) \geq E(Y)$, valendo a igualdade se, e somente se, $P(X = Y) = 1$

Proposição 21. A esperança das distribuições vistas são

1. $X \sim \text{Bern}(p) \Rightarrow E(X) = p$
2. $X \sim \text{Bin}(n, p) \Rightarrow E(X) = np$
3. $X \sim \text{HGeom}(m, n, p) \Rightarrow E(X) = \dfrac{tm}{m + n}$
4. $X \sim \text{DUnif}(C) \Rightarrow E(X) = \dfrac{\sum_{x \in \text{Supp}(X)} x}{|C|}$
5. $X \sim \text{Geom}(p) \Rightarrow E(X) = \dfrac{1 - p}{p}$
6. $X \sim \text{FS}(p) \Rightarrow E(X) = \dfrac{1}{p}$
7. $X \sim \text{NBin}(n, p) \Rightarrow E(X) = \dfrac{n(1 - p)}{p}$
8. $X \sim \text{Pois}(\lambda) \Rightarrow E(X) = \lambda$

Variáveis Aleatórias Indicadoras

Definição 17. Seja $A \subset S$. Então, a variável aleatória indicadora de A é a função $I_A : S \to \mathbb{R}$ dada por

$$I_A(x) = \begin{cases} 1 & \text{se } x \in A \\ 0 & \text{se } x \notin A \end{cases}$$

Proposição 22. Sejam $A, B \subset S$. Então

1. $(I_A)^k = I_A$
2. $I_{A^C} = 1 - I_A$
3. $I_{A \cap B} = I_A I_B$
4. $I_{A \cup B} = I_A + I_B - I_{A \cap B}$

Proposição 23 (A ponte Fundamental)**.** $P(A) = E(I_A)$

Proposição 24 (Desigualdades de Bonferroni)**.** Se k é ímpar, então

$$I\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{j=1}^{k} \left[(-1)^{j+1} \sum_{1 \leq i_1 < \ldots < i_j \leq n} I(A_{i_1} \cap \ldots \cap A_{i_j})\right]$$

Se k é par, então

$$I\left(\bigcup_{i=1}^{n} A_i\right) \geq \sum_{j=1}^{k} \left[(-1)^{j+1} \sum_{1 \leq i_1 < \ldots < i_j \leq n} I(A_{i_1} \cap \ldots \cap A_{i_j})\right]$$

O mesmo ocorre com probabilidades, em virtude da ponte fundamental.

Proposição 25. Seja X uma variável aleatória não-negativa. Então

$$E(X) = \sum_{n=0}^{\infty} (1 - F(x))$$

(LOTUS) Lei do Estatístico Inconsciente

Proposição 26. Seja X uma variável aleatória discreta e $g : \mathbb{R} \to \mathbb{R}$ uma função. Então

$$E(g(X)) = \sum_{x \in \text{Supp}(X)} g(x) P(X = x)$$

Variância

Definição 18. A variância de uma variável aleatória X é definida como

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

Comentário: A variância pode ser entendida como uma "medida" de o quanto dispersa está a distribuição de X em torno de sua média $E(X)$.

Definição 19. Se X é uma variável aleatória, então seu desvio padrão é dado por

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

Proposição 27. Para qualquer variável aleatória X, tem-se

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

Proposição 28. A variância de algumas das distribuições vistas são

1. $X \sim \text{Bern}(p) \Rightarrow \text{Var}(X) = p(1-p)$
2. $X \sim \text{Bin}(n,p) \Rightarrow \text{Var}(X) = np(1-p)$
3. $X \sim \text{Geom}(p) \Rightarrow \text{Var}(X) = \dfrac{1-p}{p^2}$
4. $X \sim \text{NBin}(n,p) \Rightarrow \text{Var}(X) = \dfrac{n(1-p)}{p^2}$
5. $X \sim \text{Pois}(\lambda) \Rightarrow \text{Var}(X) = \lambda$

Variáveis Aleatórias Contínuas

Introdução

Definição 20. Uma variável aleatória é dita contínua se sua função de distribuição de probabilidade (CDF) $F_X(x) = P(X \leq x)$ é diferenciável, exceto possivelmente em alguns pontos.

Nota: Esta definição engloba o caso de variáveis aleatórias discretas, que apresentam saltos em pontos de sua CDF e derivada 0 fora desses pontos. Porem, não convém utilizar a derivada destas funções neste caso. Por isso, vamos considerar variáveis discretas e contínuas como objetos distintos. Daremos adiante uma definição alternativa que resolve este problema.

Definição 21. A função de densidade de probabilidade (PDF) f_X de uma variável aleatória contínua X é

$$f_X(x) = \frac{d}{dx}F_X(x) = F'_X(x)$$

Nota: Nos pontos onde $F_X(x)$ não é derivável, esta definição não funciona. Por isso, a escolha dos valores de f_X nestes pontos é arbitrária, visto que não vão causar impacto nas contas. Além disso, f_X só irá existir se F_X não tiver "saltos", por isso convencionamos que ao tratar de variáveis contínuas, não haverá "saltos".

Proposição 29. A CDF de uma variável contínua X é dada por

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt$$

Nota: Segue disto que $P(a \leq X \leq b) = \int\limits_a^b f(t)dt$ e também que $P(X = 0) = 0$.

Proposição 30. A PDF f_X de uma variável contínua X satisfaz

- $f_X \geq 0$
- $\int\limits_{-\infty}^{\infty} f_X(t)dt = 1$

Reciprocamente, uma função f que satisfaz estas duas condições é a PDF de alguma variável contínua X.

Nota: Isto nos permite poder definir uma variável contínua X pela existência de uma função $f_X \geq 0$ integrável tal que sua CDF é dada por

$$F_X(x) = \int_{-\infty}^{x} f(t)dt$$

Daí, segue que $\int\limits_{-\infty}^{\infty} f(t)dt = 1$, que não seria possível para variáveis discretas, cuja integral daria 0 (pois a derivada seria 0 em quase todo ponto)

Definição 22. A esperança de uma variável contínua X (se existir) é

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

Proposição 31 (LOTUS)**.** A esperança de $F(X)$, onde X é uma variável contínua e F é uma função, é

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

Proposição 32. A esperança de uma variável contínua não negativa X é dada por

$$E(X) = \int_0^{\infty} G(x)dx$$

onde $G(x) = P(X > x) = 1 - F_X(x)$ é a função de sobrevivência de X.

Nota: De modo geral, para qualquer variável contínua X, temos

$$E(X) = -\int_{-\infty}^{0} F_X(x)dx + \int_0^{\infty} G(x)dx$$

Proposição 33. Sejam $X_1, ..., X_n$ distribuições contínuas i.i.d. Então, para toda permutação $(a_1, ..., a_n)$ de $(1, ..., n)$, tem-se $P(X_{a_1} < ... < X_{a_n}) = \dfrac{1}{n!}$.

Transformações de Localização e de Escala

Definição 23. Seja X uma variável contínua e $Y = \sigma X + \mu$, com $\sigma > 0$. Então, Y é uma transformação de localização e escala de X. O parâmetro μ muda a localização e σ muda a escala.

Nota: Essa definição não será de grande interesse, mas sim a ideia de tomar variáveis da forma $Y = \sigma X + \mu$, a fim de facilitar as contas, pois, por exemplo, conhecendo-se EX e $\text{Var}(X)$, conhece-se também $EY = E(\sigma X + \mu) = \sigma EX + \mu$ e conhecendo-se $\text{Var}(X)$, conhece-se também $\text{Var}(X) = \text{Var}(\sigma X + \mu) = \sigma^2 \text{Var}(X)$.

O Processo de Poisson

"Considere o número de ocorrências de um certo fenômeno no decorrer do tempo, por exemplo, o número de telefonemas que chegam em uma central telefônica a uma taxa média λ. Contamos o número de telefonemas que chegam até o tempo $t \geq 0$ em uma variável aleatória X_t". Sob certas condições (bastante razoáveis), obtemos que $X_t \sim \text{Pois}(\lambda t)$. Tais condições são:

- **incrementos estacionários:** A probabilidade de chegada de telefonemas no intervalo $(s, s+t]$ depende somente de t e não de s, de modo que para o cálculo das probabilidades, necessitamos somente do número de ligações que ocorrem no intervalo $[0, t)$.

- **incrementos independentes:** Os números de chegadas durante intervalos disjuntos de tempo são independentes, isto é, não há qualquer relação entre o número de ligações em momentos distintos.

- **as chamadas chegam sozinhas e não simultaneamente:** Podemos interpretar isto, em termo de probabilidades condicionais, do seguinte modo

$$\lim_{t \to 0} P(\text{chegar } n > k \text{ ligações em } (0, t] | \text{chegaram } k \text{ ligações em } (0, t]) = 0$$

Sendo T_1 a variável aleatória que mede o tempo até a primeira ligação, temos que $T_1 \sim \text{Expo}(\lambda)$ e, também, $T_2 - T_1 \sim \text{Expo}(\lambda)$. De modo geral, $T_{n+1} - T_n \sim \text{Expo}(\lambda)$, onde T_i é a variável aleatória que mede o tempo até a i-ésima ligação. Além disso, $T_2, T_3, \ldots$ não são exponenciais, mas são gammas (estas distribuições serão vistas a seguir). Em resumo,

- O número de ocorrências em $[0, t)$ é distribuído por $\text{Pois}(\lambda t)$.
- O tempo entre duas ocorrências consecutivas é distribuído por $\text{Expo}(\lambda)$.

Algumas Distribuições Contínuas

Nota: Em alguns casos, vamos definir a distribuição apenas pela função de densidade pelo fato da distribuição acumulada (integral da densidade) não adimitir uma expressão em termos elementares

- **Uniforme**
  - **Intuição:** Todos os valores em um intervalo (ou região) específico são igualmente prováveis.
  - **Definição:** Para um intervalo $[a, b]$, a PDF será $f_X(x) = \begin{cases} \frac{1}{b-a} & \text{se } a \leq x \leq b \\ 0 & \text{caso contrário} \end{cases}$

    Neste caso, sua CDF será $F_X(x) = \begin{cases} 0 & \text{se } x < a \\ \frac{x-a}{b-a} & \text{se } a \leq x \leq b \\ 1 & \text{se } x > b \end{cases}$
  - **Notação:** $X \sim \text{Unif}(a, b)$
  - **Propriedades:**

    (i) $E(X) = \dfrac{a+b}{2}$

    (ii) $\text{Var}(X) = \dfrac{(b-a)^2}{12}$

    Nota: Se $X \sim \text{Unif}(0, 1)$ e $Y = (b-a)X + a$, então $Y \sim \text{Unif}(a, b)$.

- **Normal**
  - **Intuição:** Representa fenômenos que tendem a agrupar-se em torno de uma média e, com isso, terá papel fundamental no chamado **Teorema Central do Limite**
  - **Definição:** A PDF será $f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (x \in \mathbb{R})$

    Neste caso, sua CDF não tem expressão em termos elementares.
  - **Notação:** $X \sim \mathcal{N}(\mu, \sigma^2)$

- **Propriedades:**
  (i) $E(X) = \mu$
  (ii) $\text{Var}(X) = \sigma^2$
  (iii) **simetria em relação à média:**
    - $f_X(\mu + x) = f_X(\mu - x)$
    - $F_X(\mu + x) = 1 - F_X(\mu - x)$
    - $X \sim \mathcal{N}(\mu, \sigma^2) \Rightarrow -X \sim \mathcal{N}(-\mu, \sigma^2)$

  Nota: Se $X \sim \mathcal{N}(0, 1)$, que é chamada de distribuição normal-padrão, e $Y = \sigma X + \mu$, então $Y \sim \mathcal{N}(\mu, \sigma^2)$. A PDF para X, neste caso, será denotada por ϕ e a CDF por Φ. Além disso, vale também a regra dos $68 - 95 - 99.7\%$, que é:
  - $P(|Y - \mu| < \sigma) \approx 0.68$
  - $P(|Y - \mu| < 2\sigma) \approx 0.95$
  - $P(|Y - \mu| < 3\sigma) \approx 0.997$

- **Exponencial**
  - **Intuição:** Tempo de espera até a primeira chegada de um Processo de Poisson.
  - **Definição:** A PDF será $f(x) = \lambda e^{-\lambda x} \quad (x > 0)$
    Neste caso, sua CDF será $F_X(x) = 1 - e^{-\lambda x} \quad (x > 0)$
  - **Notação:** $X \sim \text{Expo}(\lambda)$
  - **Propriedades:**
    (i) $E(X) = \dfrac{1}{\lambda}$
    (ii) $\text{Var}(X) = \dfrac{1}{\lambda^2}$
    (iii) (Falta de Memória) $P(X \geq s + t | X \geq s) = P(X \geq t), \forall s, t \geq 0$

    Nota: Se $X \sim \text{Expo}(1)$ e $Y = \dfrac{1}{\lambda} X$, então $Y \sim \text{Expo}(\lambda)$.

    Nota: Reciprocamente a propriedade (iii), toda variável contínua não negativa com a propriedade da falta de memória é uma exponencial.

    Nota: A falta de memória tem sua versão para o caso discreto e é

$$P(X \geq j + k | X \geq k) = P(X \geq t), \forall j, k \in \mathbb{Z}_+$$

    E, analogamente à propriedade (iii) e a Nota anterior, temos que a distribuição Geométrica satisfaz a propriedade e é a única dentre as discretas não-negativas que satisfazem.

- **Beta**

  - **Intuição:** Generalização da uniforme.
  - **Definição:** A PDF será

$$f(x) = \frac{1}{\beta(a,b)} x^{a-1}(1-x)^{b-1} \quad (0 < x < 1),$$

  onde a constante $\beta(a,b)$ é tomada de modo que a PDF integre 1.
  Neste caso, sua CDF não tem expressão em termos elementares.

  - **Notação:** $X \sim \text{Beta}(a,b)$
  - **Propriedades:**

    (i) $\beta(a,b) = \dfrac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$

    (ii) $E(X) = \dfrac{a}{a+b}$

    (iii) Se $a = b$, a PDF é simétrica em relação a $1/2$. Se $a > b$, a PDF é concentrada a direita de $1/2$; se $a < b$ a PDF é concentrada a esquerda de $1/2$.

    (iv) A distribuição beta é conjugada com a binomial, isto é, se uma parâmetro p tem distribuição a prior dado por uma beta e uma variável X, condicionada em p, tem distribuição binomial, então a distribuição posterior de p continua sendo uma beta.

  Nota: Quando $a = b = 1$, as distribuições $\text{Beta}(1,1)$ e $\text{Unif}(0,1)$ são a mesma distribuição.

- **Gamma**

  - **Intuição:** Generalização do tempo de espera até a n-ésima chegada de um processo de Poisson (Exponencial).
  - **Definição:** A PDF será

$$f(x) = \frac{1}{\Gamma(a)} (\lambda x)^a e^{-\lambda x} \frac{1}{x} \quad (x > 0),$$

  onde $\Gamma(a)$ é a função gamma avaliada em a (que é um valor tal que a PDF integre 1).
  Neste caso, sua CDF não tem expressão em termos elementares.

  - **Notação:** $X \sim \text{Gamma}(a, \lambda)$
  - **Propriedades:**

    (i) $E(X) = \dfrac{a}{\lambda}$

    (ii) $\text{Var}(X) = \dfrac{a}{\lambda^2}$

(iii) Seja $X_1, X_2 \cdots, X_n$ variáveis alteatórias i.i.d de distribuição Expo(λ). Então,

$$X_1 + \cdots + X_n \sim \text{Gamma}(n, \lambda)$$

(iv) A distribuição gamma é conjugada com a poisson, isto é, se um parâmetro p tem distribuição a prior dada por uma gamma e uma variável X, condicionada a p, tem distribuição de poisson, então a distribuição posterior de p continua sendo uma gamma.

Nota: Quando $a = 1$, as distribuições Gamma$(1, \lambda)$ e Expo(λ) são as mesmas

Algumas Distribuições Conjuntas

- **Multinomial**

  - **Intuição:** Alocação de n objetos em k categorias onde cada alocação é feita de forma independente, com probabilidade $p_j > 0$ para a j-ésima categoria.

  - **Definição:** Seja X_j o número de objetos alocados na j-ésima categoria. Então o vetor $\boldsymbol{X} = (X_1, ..., X_k)$ é dito ter distribuição multinomial de parâmetros k, n e $\boldsymbol{p} = (p_1, ..., p_k)$. Sua PMF conjunta é

$$P(X_1 = n_1, ..., X_k = n_k) = \frac{n!}{n_1! n_2! \cdots n_k!} \cdot p_1^{n_1} p_2^{n_2} \cdots p_k^{n_k}$$

  - **Notação:** $\boldsymbol{X} \sim \text{Mult}_k(n, \boldsymbol{p})$

  - **Propriedades:**

    (i) $X_j \sim \text{Bin}(n, p_j)$, para $j = 1, ..., k$

    (ii) $X_i + X_j \sim \text{Bin}(n, p_i + p_j)$, para $i \neq j$

    (iii) $(X_1 + X_2, X_3, ..., X_n) \sim \text{Mult}_{k-1}(n, (p_1 + p_2, p_3, ..., p_n))$

    (iv) $(X_2, ..., X_k)|X_1 = n_1 \sim \text{Mult}_{k-1}(n - n_1, (p'_2, ..., p'_k))$, onde $p'_j = \dfrac{p_j}{p_2 + ... + p_k}$

    (v) $\text{Cov}(X_i, X_j) = -np_i p_j$, para $i \neq j$

- **Multivariada (MVN)**

  - **Intuição:** Conjunto de variáveis aleatórias que estão correlacionadas e distribuídas normalmente.

  - **Definição:** $\mathbf{X} = (X_1, ..., X_n)$ é tal que qualquer combinação linear de $X_1, ..., X_n$ tem distribuição normal, ou seja, $t_1 X_1 + ... t_n X_n$ tem distribuição normal, $\forall t_1, ..., t_n \in \mathbb{R}$.

  - **Notação:** MVN

  - **Propriedades:**

(i) Se $\mathbf{X} = (X_1, ..., X_n)$ é uma MVN, então qualquer subvetor é uma MVN. Por exemplo, se (X_1, X_2, X_3) é uma MVN, então (X_1, X_2) é uma MVN. Em particular, X_i tem distribuição Normal, $\forall i$.

(ii) Se $\mathbf{X} = (X_1, ..., X_n)$ e $\mathbf{Y} = (Y_1, ..., Y_m)$ são MVN's e X é independente de Y, então a concatenação $(\mathbf{X}, \mathbf{Y}) = (X_1, ..., X_n, Y_1, ..., Y_m)$ é uma MVN.

Nota: Uma MVN $(X_1, ..., X_n)$ é completamente determinada pela média e a variância das componentes X_i e pela covariância entre elas $\text{Cov}(X_i, X_j)$. Por exemplo, um vetor (X, Y) com distribuição normal bivariada é completamente determinado por $E(X), E(Y), \text{Var}(X), \text{Var}(Y)$ e $\text{Cov}(X, Y)$.
Nota: Para $n = 2$, temos uma normal bivariada, cuja PDF conjunta é dada por

$$f_{X,Y}(x, y) = \frac{1}{2\pi\tau} e^{-(x^2+y^2-2\rho xy)/(2\tau^2)}$$

onde $\rho = \text{Corr}(X, Y)$ é tal que $\rho \in (-1, 1)$ e $\tau = \sqrt{1-\rho^2}$.

A Universalidade da Uniforme

Proposição 34. Seja F uma função estritamente crescente e que satisfaz as condições de uma CDF e seja $U \sim \text{Unif}(0, 1)$. Então

a) $X = F^{-1}(U)$ é uma variável contínua tal que $X \sim F_X = F$.

b) Se X é uma variável contínua tal que $F_X = F$, então $F(X) \sim \text{Unif}(0, 1)$.

Nota: Esta Proposição dá um modo prático de gerar variáveis aleatórias contínuas através de sua distribuição.

Exemplo 0.1. A distribuição logística tem CDF

$$F(x) = \frac{e^x}{1+e^x}$$

Para gerar uma variável aleatória X com esta distribuição, tomemos uma variável $U \sim \text{Unif}(0, 1)$. Agora, vamos determinar F^{-1}. Seja y tal que $F(x) = y$, ou seja, $\frac{e^x}{1+e^x} = y$. Denotando e^x por z, temos $\frac{z}{1+z} = y \Leftrightarrow z = y(1+z) \Leftrightarrow z = \frac{y}{1-y}$. Assim, $x = \ln\left(\frac{y}{1-y}\right)$. Portanto, $F^{-1}(x) = \ln\left(\frac{x}{1-x}\right)$. Portanto, basta tomar $X = \ln\left(\frac{U}{1-U}\right)$.

Estatísticas de Ordem

Definição 24. Para variáveis aleatórias X_1, X_2, ...,X_n, as estatísticas de ordem são as variáveis aleatórias $X_{(1)}$, $X_{(2)}$, .., $X_{(n)}$, onde

$$
\begin{aligned}
&X_{(1)} = \min(X_1, ..., X_n)\\
&X_{(2)} \quad \text{é a segunda menor das } X_1, ..., X_n\\
&\vdots\\
&X_{(n-1)} \text{ é a segunda maior das } X-1, ..., X_n\\
&X_{(n)} = \max(X_1, ..., X_n)
\end{aligned}
$$

Nota: Observe que $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$. Chamamos $X_{(j)}$ de j-ésima estatística de ordem. Se n é ímpar, $X_{((n+1)/2)}$ é a mediana amostral de X_1, ...,X_n. Cada uma das estatísticas de ordem é uma função de X_1, ..., X_n. Além disso, claramente as estatísticas de ordem são dependentes.

Proposição 35. Sejam X_1, ..., X_n variáveis aleatórias contínuas i.i.d com CDF F. Então a CDF da j-ésima estatística de ordem $X_{(j)}$ é

$$P(X_{(j)} \leq x) = \sum_{k=j}^{n} \binom{n}{k} F(x)^k (1 - F(x))^{n-k}$$

Proposição 36. Sejam X_1, ..., X_n variáveis aleatórias contínuas i.i.d com PDF f. Então a PDF da j-ésima estatística de ordem $X_{(j)}$ é

$$f_{(X_{(j)})}(x) = n\binom{n-1}{j-1} f(x)F(x)^{j-1}(1 - F(x))^{n-j}$$

Momentos

Introdução

Definição 25. Um valor c é dito mediana de uma variável X se $P(X \leq c) \geq 1/2$ e $P(X \geq c) \geq 1/2$.

Definição 26. Um valor c é dito moda de uma variável X, no caso discreto, se c maximiza a PMF de X, isto é, $P(X = c) \geq P(X = x)$ para todo $\forall x$ e, no caso contínuo, se c maximiza a PDF de X, isto é, $f_X(c) \geq f_X(x), \forall x$.

Nota: Uma váriavel aleatória pode ter multiplas medianas e modas.

Proposição 37. Seja X uma variável aleatória com média μ e mediana m. Então

- O valor c que minimiza o erro quadrático médio $E(X-c)^2$ é $c = \mu$.
- O valor c que minimiza o erro absoluto médio $E|X-c|$ é $c = m$.

Momentos

Definição 27. Seja X uma variável aleatória com média μ e variância σ^2. Para um inteiro positivo n, dizemos que:

- o n-ésimo momento de X é $E(X^n)$
- o n-ésimo momento central de X é $E\left[(X-\mu)^n\right]$
- o n-ésimo momento padronizado de X é $E\left[\left(\dfrac{X-\mu}{\sigma}\right)^n\right]$

Nota: O primeiro momento e o segundo momento central de uma variável aleatória são, respectivamente, sua média e sua variância.

Definição 28. Dizemos que uma variável X tem distribuição simétrica com respeito a um valor c se $X-c$ tem a mesma distribuição de $c-X$.

Nota: Se $E(X)$ existir, então o valor c deve ser igual ao mesmo. Com efeito,

$$E(X)-c=E(X-c)=E(c-X)=c-E(X)\Rightarrow c=E(X)$$

Proposição 38. Seja X uma variável contínua. Então X é simétrica com respeito a μ se, e somente se, $f_X(x)=f_X(2\mu-x)$ para todo x.

Proposição 39. Seja X uma variável aleatória simétrica com respeito a μ. Então para qualquer número ímpar m, o m-ésimo momento central $E(X-\mu)^m$ (se existir) é igual a 0.

Nota: A recíproca desta proposição não é verdadeira: podem existir distribuições assimétricas com todos os momentos centrais ímpares iguais a 0.

Momentos Amostrais

Definição 29. Sejam X_1, X_2, ..., X_n variáveis aleatórias i.i.d. Então

- o k-ésimo momento amostral é $M=\dfrac{\sum_{j=1}^{n} X_j^k}{n}$
- a média amostral $\overline{X}$ é o primeiro momento amostral $\overline{X}=\dfrac{\sum_{j=1}^{n} X_j}{n}$
- a variância amostral é $S^2=\dfrac{\sum_{j=1}^{n}(X_j-\overline{X})^2}{n-1}$

- o desvio padrão amostral é a raiz quadrada da variância amostral

Proposição 40. Sejam X_1, X_2, ..., X_n variáveis aleatórias i.i.d. com média μ e variância σ^2. Então

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

Além disso, a variância de $\overline{X}$ é

$$\mathrm{Var}(\overline{X}) = \frac{\sigma^2}{n}$$

Proposição 41. Sejam X_1, X_2, ..., X_n variáveis aleatórias i.i.d. com média μ e variância σ^2. Então

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

Função Geradora de Momentos (MGF)

Definição 30. A função geradora de momentos (MGF) de uma variável aleatória X (se existir) é, para algum $a > 0$, $M_X : (-a, a) \to \mathbb{R}$ dada por $M_X(t) = E(e^{tX})$, desde que a mesma seja limitada em $(-a, a)$. Caso contrário, dizemos que a MGF não existe.

Proposição 42. O n-ésimo momento de X pode ser obtido ao avaliar a n-ésima derivada da MGF em 0, isto é,

$$E(X^n) = M^{(n)}(0) = \left.\frac{d^n M}{dt^n}\right|_{t=0}$$

Definição 31. (MGF conjunta) A função geradora de momentos conjunta de um vetor aleatório $\boldsymbol{X} = (X_1, ..., X_k)$ é a função M definida como

$$M(\boldsymbol{t}) = E(e^{<\boldsymbol{t},\boldsymbol{X}>}) = E(e^{t1X_1+...+t_kX_k})$$

para $\boldsymbol{t} = (t_1, ..., t_k) \in \mathbb{R}^k$. Exigimos que a função M seja limitada em uma caixa finita $([a_1, b_1] \times ... \times [a_k, b_k])$ contendo a origem. Caso contrário, dizemos a MGF conjunta não existe.

Proposição 43. A MGF de uma variável aleatória determina sua distribução. Assim, se duas variáveis aleatórias têm a mesma MGF, elas têm a mesma distribuição.

Proposição 44. Se X e Y são variáveis independentes, então

$$M_{X+Y}(t) = M_X(t) \cdot M_Y(t)$$

Proposição 45. Se X é uma variável aleatória cuja MGF existe, então

$$M_{a+bX} = e^{at} M_X(bt)$$

Exemplos de MGF's

	MGF
Bern(p)	$p + e^t$
Geom(p)	$\dfrac{p}{1-qe^t}, \quad qe^t < 1$
Unif(a, b)	$\dfrac{e^{tb} - e^{ta}}{t(b-a)}, \quad t \neq 0$ e $M(0) = 1$
Bin(n, p)	$(pe^t + q)^n$
Nbin(r, p)	$\left(\dfrac{p}{1-qe^t}\right)^r, \quad qe^t < 1$
$\mathcal{N}(\mu, \sigma^2)$	$e^{\mu t + \frac{1}{2}\sigma^2 t^2}$
Expo(λ)	$\dfrac{\lambda}{\lambda - t}, \quad t < \lambda$
MVN	$e^{t_1 E(X_1) + \ldots + t_n E(X_n) + \mathrm{Var}(t_1 X_1 + \ldots + t_n X_n)/2}$

Função Geradora de Probabilidade (PGF)

Definição 32. A função geradora de probabilidades (PGF) de uma variável aleatória inteira e não-negativa X de PMF $p_k = P(X = k)$ é a função geradora da PMF. Por LOTUS, isto é

$$E(t^X) = \sum_{k=0}^{\infty} p_k t^k$$

Nota: A PGF converge para algum valor em $[-1, 1]$ para todo t em $[-1, 1]$ desde que $\sum_{k=0}^{\infty} p_k = 1$ e $|p_k t^k| \leq p_k$ para $|t| \leq 1$. Se ambas MGF e PGF existirem, para $t > 0$, temos

$$E(t^X) = E(e^{X \log t})$$

que é a MGF avaliada em $\log t$.

Proposição 46. Sejam X e Y variáveis aleatórias inteiras e não-negativas, com PGFs g_X e g_Y respectivamente. Suponha que $g_X(t) = g_Y(t)$ para todo t em $(-a, a)$, onde $0 < a < 1$. Então, X e Y têm a mesma distribuição, e sua PMF pode ser obtida tomando as derivadas de g_X, isto é

$$P(X = k) = P(Y = k) = \frac{g_X^{(k)}(0)}{k!}$$

Mistura de Variáveis

Distribuições Conjuntas

Definição 33. A PMF conjunta de duas variáveis aleatórias discretas X e Y é

$$p_{X,Y}(x, y) = P(X = x, Y = y)$$

A CDF conjunta de duas variáveis aleatórias X e Y é

$$F_{X,Y}(x, y) = P(X \leq x, Y \leq y)$$

A PDF conjunta de duas variáveis aleatórias contínuas X e Y, é

$$f_{X,Y}(x, y) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x, y)$$

Distribuições Marginais

Definição 34. A PMF marginal de X, dadas as variáveis aleatórias discretas X e Y, é

$$P(X = x) = \sum_y P(X = x, Y = y)$$

A CDF marginal de X, dadas as variáveis aleatórias X e Y, é

$$F_X(x) = P(X = x) = \lim_{y \to +\infty} P(X \leq x, Y \leq y) = \lim_{y \to +\infty} F_{X,Y}(x, y)$$

A PDF marginal de X, dadas as variáveis aleatórias contínuas X e y, é

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dy$$

Distribuições Condicionais

Definição 35. A PMF condicional de Y dado $X = x$, onde X e Y são variáveis aleatórias discretas é

$$P(Y = y|X = x) = \frac{P(X = x, Y = y)}{P(X = x)}$$

A PDF condicional de Y dado $X = x$, onde X e Y são variáveis aleatórias contínuas é

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x, y)}{f_X(x)}$$

Proposição 47. Se X, Y são variáveis aleatórias contínuas, então

$$f_{Y|X}(y|x) = \frac{f_{X|Y}(x|y)f_Y(y)}{f_X(x)}$$

Nota: A fórmula de Bayes continua valendo para uma mistura dos tipos das variáveis. Por exemplo, se X é contínua e Y é discreta, então

$$P(Y = y|X = x) = \frac{f_X(x|Y = y)P(Y = y)}{f_X(x)}$$

E se X é discreta e Y é contínua, então

$$f_Y(y|X = x) = \frac{P(Y = y|X = x)f_Y(y)}{P(X = x)}$$

Proposição 48. Se X, Y são variáveis aleatórias, com Y discreta, então

$$P(X \leq x) = \sum_y P(X \leq x|Y = y)P(Y = y)$$

Se X, Y são variáveis aleatórias, com Y contínua, então

$$P(X \leq x) = \int_{-\infty}^{\infty} P(X \leq x|Y = y)f_Y(y)dy$$

Nota: A Lei da Probabilidade Total continua valendo para uma mistura dos tipos das variáveis. Por exemplo, se X é contínua e Y é discreta, então

$$P(X \leq x) = \sum_y P(X \leq x|X = x)P(X = x)$$

E se X é discreta e Y é contínua, então

$$P(X = x) = \int_{-\infty}^{\infty} P(X = x|Y = y)f_Y(y)dy$$

Independência

Definição 36. Variáveis aleatórias X e Y são independentes se

$$F_{X,Y}(x,y) = F_X(x)F_Y(y)$$

No caso discreto, isto equivale a

$$P(X = x, Y = y) = P(X = x)P(Y = y)$$

No caso contínuo, isto equivale a

$$f_{X,Y}(x,y) = f_X(x)f_Y(y)$$

Proposição 49. Se a PMF conjunta de X, Y (discretas) fatora em funções não negativas g e h do seguinte modo

$$p_{X,Y}(x,y) = g(x)h(y)$$

Então, X e Y são independentes.
Se a CDF conjunta de X, Y fatora em funções não negativas g e h do seguinte modo

$$F_{X,Y}(x,y) = g(x)h(y)$$

Então, X e Y são independentes.
Se a PDF conjunta de X, Y (contínuas) fatora em funções não negativas g e h do seguinte modo

$$f_{X,Y}(x,y) = g(x)h(y)$$

Então, X e Y são independentes.

Seguem abaixos dois resultados importantes relativos a variáveis aleatórias independentes.

Proposição 50. $X \sim \text{Pois}(\lambda p)$ e $Y \sim \text{Pois}(\lambda(1-p))$ e X e Y são independentes se, e somente se, $X + Y \sim \text{Pois}(\lambda)$ e $X|X + Y = n \sim \text{Bin}(n, p)$.

Proposição 51. Sejam $X \sim \text{Expo}(\lambda_1)$ e $Y \sim \text{Expo}(\lambda_2)$ independentes. Então

$$P(X < Y) = \frac{\lambda_1}{\lambda_1 + \lambda_2}$$

LOTUS 2D

Proposição 52. Seja $g : \mathbb{R}^2 \to \mathbb{R}$. Então, se X e Y são discretas, então

$$E(g(X,Y)) = \sum_x \sum_y g(x,y)P(X = x, Y = y)$$

Se X e Y são contínuas, então

$$E(g(X,Y)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y)dydx$$

Nota: A ideia pode ser generalizada para n dimensões.

Covariância

Definição 37. A Covariância entre as variáveis aleatórias X e Y é

$$\text{Cov}(X, Y) = E((X - EX)(Y - EY)) = E(XY) - E(X)E(Y)$$

Proposição 53. Temos as seguintes propriedades da covariância

- $\text{Cov}(X, X) = \text{Var}(X)$
- $\text{Cov}(X, Y) = \text{Cov}(Y, X)$
- $\text{Cov}(X, c) = 0, \forall c \in \mathbb{R}$
- $\text{Cov}(aX, Y) = a\text{Cov}(X, Y)$
- $\text{Cov}(X + Y, Z) = \text{Cov}(X, Z) + \text{Cov}(Y, Z)$
- Var(X, Y) = Var(X) + Var(Y) - 2Cov(X, Y)
- Se X e Y são independentes, então $\text{Cov}(X, Y) = 0$ (A recíproca não é necessariamente verdade)

Definição 38. A correlação entre X e Y é

$$\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}$$

Proposição 54. Se X e X são variáveis aleatórias, então

$$-1 \leq \text{Corr}(X, Y) \leq 1$$

Tranformações

Mudanças de variáveis

Proposição 55. Seja X uma variável aleatória contínua com PDF f_X, e seja $Y = g(X)$, onde g é diferenciável e estritamente crescente (ou estritamente decrescente). Então a PDF de Y é dada por

$$f_Y(y) = f_X(x)\left|\frac{dx}{dy}\right|$$

onde $x = g^{-1}(y)$. O suporte de Y são todos os $g(x)$ com x no suporte de X.

Proposição 56. Seja $\boldsymbol{X} = (X_1, \cdots, X_n)$ um vetor aleatório contínuo com PDF conjunta $f_{\boldsymbol{X}}$. Seja $g : A_0 \to B_0$ uma função invertível, onde A_0 e B_0 são conjuntos abertos de $\mathbb{R}^n$, A_0 contém o suporte de $\boldsymbol{X}$, e B_0 a imagem de g.

Seja $\boldsymbol{Y} = g(\boldsymbol{X})$ e, de forma semelhante, $\boldsymbol{y} = g(\boldsymbol{x})$. Como g é invertível, também temos $\boldsymbol{X} = g^{-1}(\boldsymbol{Y})$ e $\boldsymbol{x} = g^{-1}(\boldsymbol{y})$.

Suponha que todas as derivadas parciais $\partial x_i / \partial y_j$ existam e sejam contínuas, então a matriz Jacobiana é formada como

$$\frac{\partial \boldsymbol{x}}{\partial \boldsymbol{y}} = \begin{pmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} & \cdots & \frac{\partial x_1}{\partial y_n} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} & \cdots & \frac{\partial x_2}{\partial y_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial x_n}{\partial y_1} & \frac{\partial x_n}{\partial y_2} & \cdots & \frac{\partial x_n}{\partial y_n} \end{pmatrix}$$

Assuma também que o determinante da Jacobiana nunca é 0. Então a PDF conjunta de $\boldsymbol{Y}$ é

$$f_{\boldsymbol{Y}} = f_{\boldsymbol{X}}(g^{-1}(\boldsymbol{y})) \cdot \left| \det \frac{\partial \boldsymbol{x}}{\partial \boldsymbol{y}} \right|$$

para $\boldsymbol{y} \in B_0$ e 0 caso contrário.

Covoluções

Definição 39. Uma convolução é a soma de variáveis aleatórias independentes.

Proposição 57. Sejam X e Y variáveis aleatórias independentes e $T = X + Y$ sua soma. Se X e Y são discretas, a PMF de T é

$$P(T = t) = \sum_x P(Y = t - x) P(X = x) = \sum_y P(X = t - y) P(Y = y)$$

Se X e Y são contínuas, a PDF de T é

$$f_T(t) = \int_{-\infty}^{\infty} f_Y(t - x) f_X(x) dx = \int_{-\infty}^{\infty} f_X(t - y) f_Y(y) dy$$

Esperança Condicional

Esperança dada um Evento

Definição 40. Seja A um evento. Se Y é uma variável aleatória discreta, então a esperança condicional de Y dado A é

$$E(Y|A) = \sum_y y P(Y = y|A)$$

Se Y é uma variável aleatória contínua, então a esperança condicional de Y dado A é

$$E(Y|A) = \int_{-\infty}^{\infty} y f_Y(y|A) dy$$

Nota: Uma vez que $X = x$ é um evento, $E(Y|X = x)$ é simplesmente a esperança de Y condicionada a este evento. Ou seja, se Y é discreta, temos que

$$E(Y|X = x) = \sum_y y P(Y = y|X = x)$$

e, se Y for contínua,

$$E(Y|X = x) = \int_{-\infty}^{\infty} y f_{Y|X}(y|x) dy$$

Note que, neste caso, a esperança é um número.

Proposição 58. (Lei da esperança total) Sejam $A_1, A_2, \cdots, A_n$ uma partição do espaço amostral, com $P(A_i) > 0$ para todo i, e Y uma variável aleatória. Então

$$E(Y) = \sum_{i=1}^{n} E(Y|A_i)P(A_i)$$

Esperança dada uma Variável

Definição 41. Seja $g(x) = E(Y|X = x)$. Definimos a esperança condicional de Y dado X como a variável aleatória $g(X)$ e a denotamos por $E(Y|X)$. Em outras palavras, se a realização do experimento X retorna x, então $E(Y|X)$ retorna $g(x)$.
Nota: Note que, neste caso, a esperança é uma variável aleatória.

Proposição 59. Temos as seguintes propriedades da esperança condicional

- Se X e Y são independentes, então $E(Y|X) = E(Y)$
- Para qualquer função h, $E(h(X)Y|X) = h(X)E(Y|X)$
- $E(cY_1 + Y_2|X) = cE(Y_1|X) + E(Y_2|X)$
- (Lei de Adão) $E(E(Y|X)) = E(Y)$
- (Lei de Adão com condicionamento extra) Para quaisquer variáveis aleatórias X, Y, Z, vale
$$E(E(Y|X, Z)|Z) = E(Y|Z)$$
- (Interpretação de projeção) A variável aleatória $Y - E(Y|X)$ é não correlacionada com $h(X)$, para toda função h. Equivalentemente
$$E((Y - E(Y|X))h(X)) = 0$$

Nota: $Y - E(Y|X)$ é chamada de residual em usar X para predizer Y.

Variância condicional

Definição 42. A variância condicional de Y dado X é

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

que é equivalente a

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

Proposição 60. (Lei de Eva) Para quaisquer variáveis aleatórias X e Y, vale

$$\text{Var}(Y) = E(\text{Var}(Y|X)) + \text{Var}(E(Y|X))$$

Grandes Teoremas

Desigualdades

- (Cauchy-Schwarz) Se X e Y são variáveis aleatórias com variâncias finitas, então

$$E(XY) \leq \sqrt{E(X^2)E(Y^2)}$$

- (Jensen) Se X é uma variável aleatória e g é uma função convexa, então

$$E(g(X)) \geq g(E(X))$$

  Se g é côncava, então

$$E(g(X)) \leq g(E(X))$$

  A igualdade (em ambos os casos) ocorre se $g = a + bX$, para alguns $a, b \in \mathbb{R}$.

- (Markov) Se X é uma variável aleatória e $a > 0$, então

$$P(|X| \geq a) \leq \frac{E(|X|)}{a}$$

- (Tchebychev) Se X é uma variável aleatória, com média μ e variância σ^2 e $a > 0$, então

$$P(|X - \mu| \geq a) \leq \frac{\sigma^2}{a^2}$$

- (Chernoff) Se X é uma variável aleatória e $a, t > 0$, então

$$P(X \geq a) \leq \frac{E(e^{tX})}{e^{ta}}$$

Lei dos Grandes Números

Proposição 61 (Lei Fraca dos Grandes Números)**.** Sejam $X_1, X_2, \ldots$ variáveis aleatórias i.i.d., com média μ. Então, a média amostral converge em probabilidade para μ, isto é, dado $\varepsilon > 0$, temos que

$$P\left(\left|\frac{X_1 + \ldots + X_n}{n} - \mu\right| \geq \varepsilon\right) \to 0$$

Proposição 62 (Lei Forte dos Grandes Números)**.** Sejam $X_1, X_2, \ldots$ variáveis aleatórias i.i.d., com média μ. Então, a média amostral converge quase certamente para μ, isto é

$$P\left(\frac{X_1 + \ldots + X_n}{n} \to \mu\right) = 1$$

O Teorema Central do Limite

Proposição 63 (Teorema Central do Limite)**.** Sejam $X_1, X_2, \ldots$ variáveis aleatórias i.i.d. com média μ e variância σ^2. Então, a média amostral padronizada converge em distribuição para $\mathcal{N}(0,1)$, isto é

$$\sqrt{n}\left(\frac{\frac{X_1 + \ldots + X_n}{n} - \mu}{\sigma}\right) \xrightarrow{D} \mathcal{N}(0,1)$$

Por uma transformação de localização-escala, segue que a média amostral é aproximadamente distribuída como uma $\mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$, isto é

$$\frac{X_1 + \ldots + X_n}{n} \approx \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$$