Probabilidade - MAT0234 Medida e Integração

Prof. Jorge Adrian Beloqui

1 Introdução

Em 1933, Kolmogorov publicou seu livro Fundamentos da Teoria da Probabilidade. Ele afirma que a partir dos trabalhos de Lebesgue com a medida e posteriormente com Fréchet, que generalizou a medida de Lebesgue para espaços mais gerais, foi possível axiomatizar a Teoria das Probabilidades, o que faz no livro.
Ele expressa "Depois da publicação dos trabalhos de Lebesgue, as analogias entre a medida de um conjunto e a probabilidade de um evento e entre a integral de uma função e a esperança (média) de uma variável aleatória ficaram aparentes. Estas analogias permitiram outras extensões. Por exemplo, várias propriedades de variáveis aleatórias independentes tinham similaridade completa com as propriedades correspondentes de funções ortogonais. Mas se a Teoria da Probabilidade devesse se embasar nestas analogias, seria necessário tornar as teorias da medida e de integração independentes dos elementos geométricos constituintes do contexto de Lebesgue. Isto foi realizado por Fréchet."
Vale a pena ler este Prefácio, que está na página da disciplina. Agora vamos dar os termos na área de Análise e na de Probabilidade, usados para designar objetos "análogos".

Análise	Probabilidade
espaço de medida $(X, \mathcal{A}, \mu)$, $\mu(X) = 1$	Espaço amostral $(\Omega, \mathcal{B}, p)$
(σ)-álgebra	(σ)-campo
conjunto mensurável	evento
função mensurável f	variável aleatória X
Integral de f, $\int f \, d\mu$	Esperança ou média de X, E(X)
pertencer a Lp	ter p-momento finito
convergência em medida, q.t.p	convergência em probabilidade, quase certo
função caraterística	função indicadora
Transformada de Fourier de f	função caraterística de f

De agora em diante, a menos que explicitemos o contrário, μ será uma probabilidade.
Uma noção fundamental da Probabilidade é a de **independência**. Intuitivamente, dois eventos (conjuntos) se dizem independentes se a ocorrência de um deles , não afeta a probabilidade do outro evento.
Exemplos:

1. jogar uma moeda 2 vezes

2. jogar um dado n vezes

Definição 1.1. Dados dois conjuntos A e B, $\mu(A) > 0$ a **probabilidade condicional de B dado A** é $\mu_A(B) = \mu(B|A) = \mu(A \cap B)/\mu(A)$

Definição 1.2. Dada uma sequência de conjuntos A_n. Eles chamam-se de **independentes** se $\mu(\cap_{k=1}^{m} A_{j_k}) = \prod_{k=1}^{m} \mu(A_{j_k})$

No caso de dois conjuntos A e B independentes, $\mu(A) > 0$, temos $\mu(A \cap B) = \mu(A)\mu(B)$. Neste caso, a probabilidade condicional de B dado A é

$\mu_A(B) = \mu(A)\mu(B)/\mu(A) = \mu(B)$. Ou seja, a probabilidade de ocorrência do evento B não é afetada pela ocorrência do evento A.
Exemplos:

1. seja uma caixa com 5 bolas brancas e cinco bolas pretas. Escolhamos 2 bolas consecutivamente, sem substituição. Os eventos A= a primeira bola é branca, B= a segunda bola é preta são independentes?

2. seja uma caixa com 5 bolas brancas e cinco bolas pretas. Escolhamos 2 bolas consecutivamente, com substituição. Mostre agora que A e B são independentes.

Proposição 1. *Uma sequência de conjuntos $\{A_n\}$ é independente $\Leftrightarrow$ a sequência de conjuntos $\{A_n^c\}$ for independente.*

Solução: mostre que se $\{A_n\}$ for independente então $\{A_1^c, A_j;\ j \geq 2\}$ é independente.
Exemplo:

1. $\{1, \ldots, 9\}$ com medida $(1/9)$ para cada ponto. Os conjuntos $\{1, 2, 3\}, \{2, 4, 6\}, \{1, 4, 5\}$ são independentes dois a dois, mas não os 3 juntos.

Definição 1.3. As variáveis aleatórias (funções mensuráveis) $X_1, ..., X_n$ são **independentes** se $\mu(X_1 \in B_1, \ldots, X_n \in B_n) = \prod_1^n \mu(\{X_j \in B_j\})$ para todos conjuntos de Borel $B_1, ..., B_n$ em R.

Exemplo:

1. sejam A, B dois conjuntos independentes para μ. Logo χ_A e χ_B são independentes.

2. sejam $\varphi_1 = \sum_1^n \alpha_j \chi_{A_j}$ e $\varphi_2 = \sum_1^m \beta_k \chi_{B_k}$ onde $\{A_j\}$ e $\{B_k\}$ disjuntos 2 a 2 respectivamente. Então φ_1 independente de φ_2 se cada A_j for independente de todos os B_k e recíprocamente.

3. Seja $X = \prod_1^\infty \{1, -1\}$ onde a medida de $\{1, -1\}$ é $\mu(1) = \mu(-1) = (1/2)$. Considere a medida produto. Sejam f_j as funções $f_j(x) = x_j$ o valor da j-ésima coordenada de x. Então as f_j são funções independentes.

4. Sejam $f_j : [0, 1] \to \mathbb{R}$ definidas do seguinte modo. Se $0, x_1 x_2 \ldots x_n \ldots$ for o desenvolvimento binário de x em [0,1], tomamos $f_j(x) = x_j$. Ou seja, as imagens das f_j são 0 e 1. Estas funções são independentes.

2 Lema de Borel-Cantelli

Lema 2. *a. Seja uma sequência de eventos (conjuntos mensuráveis) $\{A_n\}$ tais que $\sum_1^\infty \mu(A_n) < +\infty$. Então $\mu(\limsup A_n) = 0$*
b. Seja uma sequência de eventos (conjuntos mensuráveis) $\{A_n\}$ independentes tais que $\sum_1^\infty \mu(A_n) = +\infty$. Então $\mu(\limsup A_n) = 1$

Demonstração. O $\limsup A_n$ é o conjunto dos x tais que pertencem a um infinidade de A_n.

a. $\limsup A_n = \cap_{k=1}^{\infty}(\cup_{n=k}^{\infty} A_n)$. Logo pela continuidade para acima de μ segue que $\mu(limsupA_n) \leq \mu(\cup_{n=k}^{\infty} A_n) \leq \sum_{n=k}^{\infty} \mu(A_n)$. Como a série $\sum_1^{\infty} \mu(A_n)$ converge, então $\lim_{k\to\infty} \sum_{n=k}^{\infty} \mu(A_n) = 0$.
b. como os A_n são independentes, então A_n^c também são independentes. Analisemos $(\limsup A_n)^c = \cup_{k=1}^{\infty}(\cap_{n=k}^{\infty} A_n^c)$. Segue que

$$\mu(\cap_{n=k}^{m} A_n^c) = \prod_{k}^{m} (1 - \mu(A_n)) \leq \prod_{k}^{m} e^{-\mu(A_n)} = e^{-\sum_{n=k}^{m} \mu(A_n)}$$

Como $\sum_{n=k}^{\infty} \mu(A_n) = \infty$ temos que $\mu(\cap_{n=k}^{\infty} A_n^c) = 0 \; \forall \; k$. Assim $\mu(\limsup A_n)^c = 0$ e $\mu(\limsup A_n) = 1$. $\square$

Aplicação: Jogamos enumeráveis vezes uma moeda equilibrada. Qual é a probabilidade de que tenhamos infinitas vezes 100 coroas consecutivas?
Consideremos $\{1, -1\}$ com a probabilidade $1/2$ para cada ponto. E o espaço amostral (espaço de medida) $X = \prod_1^{\infty} \{1, -1\}$ com a medida produto. Seja $E = \{x \in X \mid suas\ coordenadas\ tem\ 100\ coroas (= 1)\ consecutivas\ infinitas\ vezes\} = \{x \in X \mid dado\ m \in \mathbb{N}\ \exists\ n > m,\ x_{n+1} = 1, \ldots, x_{n+100} = 1\}$. Tomemos agora $A_j = \{x \in X | x_{j+1} = \cdots = x_{j+100} = 1\}$. Como μ é a medida produto, $\mu(A_j) = 1/2^{100}$. Destes A_j selecionamos a subsequência $B_j = A_{100j}$. Assim, os B_j são conjuntos independentes (exercício). Afirmamos que $\limsup\ B_j \subseteq E$. Com efeito: $C_k = \cup_{j=k}^{\infty} B_j$ é o conjunto (evento) dos pontos tais que existe $j \geq k$ com $x \in B_j$. Daí que $x \in \cap_{k=1}^{\infty} C_k \Leftrightarrow x \in C_k \; \forall \; k \geq 1 \Leftrightarrow \forall \; k \; \exists \; j,\ j \geq k, com\ x \in B_j$.
Logo, aplicando o Lema de Borel- Cantelli à sequência de conjuntos (eventos) B_j segue que $\mu(\limsup\ B_j) = 1$ e portanto $\mu(E) = 1$.

3 Algumas definições

Definição 3.1. Seja $f : X \to R$ função mensurável. Então:

1. a **média de f** é dada por $a = \int f d\mu$ quando existir.
2. a **variância de f** é dada por $\sigma^2 = \int (f - a)^2 d\mu$

   A variância é uma magnitude muito utilizada na Teoria de Probabilidade.

Exemplo: Distribuição normal $\int_{-\infty}^{x} e^{-t^2/2} dt$

1. A média da distribuição normal é zero.
2. a variância da distribuição normal é $\sqrt{2\pi}$.
3. Mostre que se $b \in \mathbb{R}$ então $\sigma^2(bf) = b^2\sigma^2(f)$. Ou seja, σ^2 é homogêneo de ordem 2.

4 Leis dos Grandes Números

Quando jogamos uma moeda equilibrada temos dois resultados possíveis, equiprováveis. Assim, atribuindo o valor 1 se sair cara e -1 se sair coroa, a média para cada jogada é 0. A ideia destas Leis é a de que quando vou jogando muitas vezes esta moeda, o número de caras e coroas tende a "ser o mesmo". Ou seja, atribuindo o valor 1 se sair cara e -1 se sair coroa, esperamos que a média tenda para zero. Observe que se atribuíssemos outros valor para cara e coroa, com média λ diferente, também esperaríamos que as médias das jogadas sucessivas convirjam para λ.

4.1 Integral de um produto de funções independentes

Tínhamos dado a definição de conjuntos independentes e de funções independentes. Como exemplo, vimos que duas funções simples φ_1; φ_2 são independentes se e somente cada conjunto da expressão padrão de uma delas é independente dos conjuntos da expressão padrão da outra.

Realizando cálculos diretos, isto nos leva a que $\int \varphi_1\varphi_2 d\mu \ = \ \int \varphi_1 d\mu \ \int \varphi_2 d\mu$. Ou seja, **a média do produto de 2 funções independentes é o produto das médias**. Isto acontece em geral para qualquer par de funções independentes f e g. Por exemplo, se a média de uma delas for zero, a integral do produto será zero, sempre que a média da outra função for finita.
Também, a existência de um dos lados da equação implica a existência do outro lado (no caso do produto, existência de cada fator). Estes resultados são simples, mas não temos tempo para demonstrá-los. Como vocês devem imaginar, as funções independentes são peculiares porque em geral o produto de duas funções de L1 não está necessariamente em L1. É o que acontece por exemplo com $f(x) = g(x) = 1/\sqrt{x}$.

4.2 Desigualdade de Kolmogorov

Inicialmente lembremos da Desigualdade de Chebyshev.

Proposição 3 (Desigualdade de Chebyshev). *Seja $f \in L^p$ e $c > 0$, então:*

$$\int |f|^p d\mu \geq c^p \ \mu(\{x \in X \mid \ |f(x)| \ \geq \ c\})$$

No campo da Probabilidade há uma versão mais corrente:
Se $a \ = \ \int f d\mu$, tomando p=2 na Desigualdade acima temos:

$$\int |f-a|^2 d\mu \geq c^2 \ \mu(\{x \in X \mid \ |f(x)-a| \ \geq \ c\})$$

Usando a notação da Probabilidade, escrevemos

$$\sigma^2(f) \ \geq c^2 \ \mu(\{x \in X \mid \ |f(x)-a| \ \geq \ c\})$$

É claro que a desigualdade vale mesmo que $\int |f|^p d\mu$ ou embaixo, $\sigma^2(f)$ sejam ∞. Mas é menos interessante. Há uma outra desigualdade que generaliza a de Chebyshev, considerando uma soma de funções mensuráveis independentes, o que abre o caminho para calcular médias.

Teorema 4 (Desigualdade de Kolmogorov). *Sejam $f_j, 1 \leq j \leq n$ funções independentes com média zero, ou seja $\int f_j \, d\mu = 0$ e $\sigma^2(f_j) < \infty$. Chamemos de $f(x) = max\{|\sum_1^k f_j(x)|, 1 \leq j \leq n\}$, ou seja, o máximo das somas parciais das f_j. Então, $\forall \, \varepsilon > 0$ temos*

$$\sum_1^n \sigma^2(f_j) \geq \varepsilon^2 \, \mu(\{x \in X \mid |f(x)| \geq \varepsilon\})$$

.

Demonstração. Seja $E = \{x \mid |f(x)| \geq \varepsilon\}$ e $s_k = \sum_1^k f_j$. Decompomos E entre os $E_k = \{|s_k(x)| \geq \varepsilon\} \cap_{j=1}^{k-1} |s_j(x)| \leq \varepsilon\}$.
Temos $\int_{E_k} s_n^2 d\mu = \int_{E_k} [(s_n - s_k)^2 + s_k^2 + 2s_k(s_n - s_k)]d\mu$. Notemos que $\int_{E_k} s_k(s_n - s_k)d\mu = \int_X \chi_{E_k} s_k(s_n - s_k)d\mu$
O fator $\chi_{E_k} s_k$ depende das funções $f_j; 1 \leq j \leq k$. Já o fator $(s_n - s_k)$ depende das $f_j; k+1 \leq j \leq n$. Portanto, $\int_X \chi_{E_k} s_k(s_n - s_k)d\mu = \int_X \chi_{E_k} s_k d\mu \int_X (s_n - s_k)d\mu$ pela independência. Como as médias das f_j são zero, o segundo fator é zero, porque $\int_X (s_n - s_k)d\mu = \int_X \sum_{k+1}^n f_j d\mu = 0$.
Logo $\int_{E_k} s_n^2 d\mu \geq \int_{E_k} s_k^2 d\mu \geq \varepsilon^2 \mu(E_k)$. Observamos que $E = \cup_1^n E_k$, onde os E_k são disjuntos dois a dois. Assim:$\sum_1^n \sigma^2(f_j) = \int (f_1 + \cdots + f_n)^2 d\mu \geq \int_E s_n^2 d\mu = \sum_1^n \int_{E_k} s_n^2 d\mu \geq \sum_1^n \mu(E_k)\varepsilon^2 = \mu(E)\varepsilon^2$. □

Observe que para o caso em que n=1, temos a Desigualdade de Chebyshev.

Teorema 5 (Teorema de Bernoulli ou Lei Fraca dos Grandes Números). *Seja f_n uma sequência de funções independentes com média zero, ou seja $\int f_j \, d\mu = 0$ e $\sigma^2(f_j) < \infty$. Se $lim_{n \to \infty} (1/n^2) \sum_1^n \sigma^2(f_j) = 0$, então a sequência das médias $\{(1/n) \sum_1^n (f_j)\}$ converge para zero em medida.*

Demonstração. Calculemos

$$\sigma^2 \left(\frac{1}{n} \sum_{j=1}^n f_j \right) = \int \left(\frac{1}{n} \sum_{j=1}^n f_j \right)^2 d\mu = \frac{1}{n^2} \int \left(\sum_{j=1}^n f_j \right)^2 d\mu = \frac{1}{n^2} \sum_{j=1}^n \sigma^2(f_j)$$

pela independência das f_j. Portanto

$$\int \left(\frac{1}{n} \sum_{j=1}^n f_j \right)^2 d\mu \to 0$$

ou seja $\frac{1}{n} \sum_{j=1}^n f_j \to 0$ em média quadrática (em L^2) e portanto converge para zero em medida □

Agora queremos ver a convergência q.t.p, para demonstrarmos a Lei Forte dos Grandes Números

Teorema 6. *Seja $\{f_n\}$ sequência de funções independentes com $\int f_n d\mu = 0$ para todo n e $\sum_{n=1}^\infty \sigma^2(f_n) < +\infty$. Então a série $\sum_{n=1}^\infty f_n(x)$ converge q.t.p.*

Demonstração. Para demonstrar usaremos a convergência em medida.
Sejam $s_n(x) = \sum_{j=1}^{n} f_j(x)$, $a_m(x) = \sup\{|s_{m+k}(x) - s_m(x)| \mid k \in \mathbb{N}\}$ e $a(x) = \inf\{a_m(x) \mid m \in \mathbb{N}\}$.
Uma condição necessária e suficiente para a convergência de $s_n(x)$ é que $a_n(x) = 0$.
Pela desigualdade de Kolmogorov temos que para todo $\varepsilon > 0$ e para todo par de naturais m e n:

$$\mu\left(\{x \mid \max_{1\le k\le n} |s_{m+k}(x) - s_m(x)| \ge \varepsilon\}\right) \le \frac{1}{\varepsilon^2} \sum_{k=m+1}^{m+n} \sigma^2(f_k)$$

segue que:

$$\mu\left(\{x \mid a_m(x) \ge \varepsilon\}\right) \le \frac{1}{\varepsilon^2} \sum_{k=m+1}^{\infty} \sigma^2(f_k)$$

e portanto $a(x) = \inf a_m(x)$ verifica:

$$\mu\left(\{x \mid a(x) \ge \varepsilon\}\right) \le \frac{1}{\varepsilon^2} \sum_{k=m+1}^{\infty} \sigma^2(f_k)$$

Como $\sum_{k=1}^{\infty} \sigma^2(f_k)$ converge, temos que $\sum_{k=m}^{\infty} \sigma^2(f_k) \to 0$ quando $m \to \infty$ e assim $\mu(\{x \mid a(x) \ge \varepsilon\}) = 0$ para todo $\varepsilon > 0$. Logo s_n converge no complementar de $A = \cup_{k=1}^{\infty}\{x \mid a(x) \ge \frac{1}{k}\}$, e $\mu(A) = 0$. Assim $s_n(x)$ converge q.t.p. □

Então vimos que se as funções tiverem média zero e a série das variâncias for convergente, então a série das funções $\sum_{n=1}^{\infty} f_n(x)$ converge q.t.p. Nada nos diz do limite; só que é finito em qtp.

Para demonstrar a Lei Forte dos Grandes Números aceitaremos o seguinte resultado:

Lema 7. *Sejam $\{y_n\}$ reais tais que $\sum_{j=1}^{\infty} \frac{y_j}{j}$ converge. Então $\lim \frac{1}{n} \sum_{j=1}^{n} y_j \to 0$.*

Teorema 8 (Lei Forte dos Grande Números)**.** *Seja $\{f_n\}$ sequência de funções independentes com variâncias finitas e média zero, ou $\int f_n d\mu = 0$. Suponhamos que $\sum_{n=1}^{\infty} \frac{\sigma^2(f_n)}{n^2} < \infty$. Então $\frac{1}{n} \sum_{k=1}^{n} f_k \to 0$ q.t.p.*

Demonstração. Definimos $g_n = \frac{f_n}{n}$. Como $\int g_n d\mu = 0$ e $\sum_{n=1}^{\infty} \sigma^2(g_n) = \sum_{n=1}^{\infty} \frac{\sigma^2(f_n)}{n^2} < \infty$ por hipótese, a séria $\sum_{n=1}^{\infty} \frac{f_n(x)}{n}$ converge q.t.p. Pelo lema anterior, $\frac{1}{n} \sum_{k=1}^{n} f_k(x) \to 0$ q.t.p. □

Uma aplicação é o teorema de Borel sobre números normais:

Teorema 9. *Quase todo ponto $x \in [0, 1]$ tem em sua expansão binária um "número igual"de 0 e 1. Isso vale para qualquer outra base.*

Demonstração. Seja $X = [0, 1]$ e consideremos os $x \in [0, 1]$ com sua expansão binária. Assim, temos todas as sequência binárias possíveis, depois da ",".

Consideremos $\{f_j\}$ sequência de funções independentes dada por: Para $x = 0, x_1 x_2 \ldots x_n \ldots$

$$f_j(x) = \begin{cases} 1 \text{ se } x_j = 1 \\ -1 \text{ se } x_j = 0 \end{cases}$$

Logo $\int f_j d\mu = (1)\frac{1}{2} + (-1)\frac{1}{2} = 0$ e $\sigma^2(f_j) = \int f_j^2 d\mu = 1$. Então

$$\sum_{n=1}^{\infty} \frac{\sigma^2(f_n)}{n^2} = \sum_{n=1}^{\infty} \frac{1}{n^2} < \infty$$

Pela lei forte dos grandes números $\frac{1}{n}\sum_{j=1}^{n} f_j(x) \to 0$. □

Vejamos agora a demonstração de dois lemas utilizados para a prova da lei forte:

Lema 10. *Se $y_n \to a \in \mathbb{R}$ então $\lim_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n} y_j = a$ (se $y_n \to a$ então a média das y_n também tende para a).*

Demonstração. Seja $\varepsilon > 0$. Então existe $n_0 \in \mathbb{N}$ com $|y_n - a| < \frac{\varepsilon}{2}$ se $n \geq n_0$. Portanto, $(1/m)\sum_{n_0}^{n_0+m} |y_n - a| < \frac{\varepsilon}{2}$ para qualquer m.
Seja agora $n_1 \in \mathbb{N}$ com $n_1 \geq n_0$ tal que $\frac{1}{n_1}\sum_{j=1}^{n_0} |y_j - a| < \frac{\varepsilon}{2}$. Tomando $n > n_1$ temos:

$$\left| \left(\frac{1}{n}\sum_{j=1}^{n} y_j \right) - a \right| = \left| \frac{1}{n}\sum_{j=1}^{n} (y_j - a) \right| \leq \frac{1}{n}\left| \sum_{j=1}^{n_0} (y_j - a) \right| + \frac{1}{n}\left| \sum_{j=n_0+1}^{n} (y_j - a) \right|$$
$$< \frac{1}{n_1}\sum_{j=1}^{n_0} |y_j - a| + \left(\frac{n - n_0}{n} \right) \frac{\varepsilon}{2} < \varepsilon$$

□

Observe que a recíproca é falsa: com efeito, tomando $y_n = (-1)^{n+1}$ temos que y_n diverge, mas $\lim_{n\to\infty} \frac{1}{n}\sum_{j=1}^{n} y_j = (1/2)$. Com efeito, as somas parciais são $S_1 = 1, S_2 = 0; S_3 = 1; S_4 = 0$ e assim seguindo. Ou seja $S_{2k+1} = 1$ e $S_{2k} = 0$.
Este outro tipo de convergência para uma sequência chama-se de "convergência Cesàro" e é utilizada por exemplo para as Séries de Fourier.

Lema 11. *Seja $\{y_n\}$ sequência de reais tal que a série $\sum_{j=1}^{\infty} \frac{y_n}{n}$ converge. Então $\lim_n \frac{1}{n}\sum_{j=1}^{n} y_n = 0$. (Observe que, somente pelas expressões das séries, nota-se que $\frac{1}{n}\sum_{j=1}^{n} y_n$ é muito menor do que $\sum_{j=1}^{\infty} \frac{y_n}{n}$ e que a segunda põe muito mais peso aos primeiros termos do que a primeira, que distribui o peso igualmente entre todos os y_n).*

Demonstração. Seja $s_n = \sum_{j=1}^{n} \frac{y_j}{j}$, $s_0 = 0$ e $t_n = \sum_{j=1}^{n} y_j$, queremos ver que $\frac{t_n}{n} \to 0$. Escrevamos y_j em termos que s_j e s_{j-1}: $y_j = j(s_j - s_{j-1})$. Logo $t_{n+1} = \sum_{j=1}^{n+1} j s_j - \sum_{j=1}^{n+1} j s_{j-1} = (n+1)s_{n+1} - \sum_{j=1}^{n} s_j$, dividindo por $n+1$ temos:

$$\frac{t_{n+1}}{n+1} = s_{n+1} - \frac{1}{n+1}\sum_{j=1}^{n} s_j = s_{n+1} - \left(\frac{n}{n+1} \right) \frac{1}{n}\sum_{j=1}^{n} s_j$$

Note que, por hipótese, s_{n+1} converge. Portanto, pelo lema anterior, $\frac{1}{n}\sum_{j=1}^{n} s_j$ converge para o mesmo valor e assim $\frac{t_{n+1}}{n+1} \to 0$.

$\square$