Kỳ vọng 2

Toán Chuyên Đề

HUST

Ngày 5 tháng 11 năm 2016

Tài liệu tham khảo

- Eric Lehman, F Thomson Leighton & Albert R Meyer, *Mathematics for Computer Science*, 2013 (Miễn phí)
- Michael Mitzenmacher và Eli Upfal, *Probability and Computing*, 2005
- Nguyễn Tiến Dũng và Đỗ Đức Thái, *Nhập Môn Hiện Đại Xác Suất & Thống Kê*.

Định lý

Với mọi biến ngẫu nhiên $R_1, R_2, \ldots, R_n$,

$$\mathrm{Ex}[R_1 + R_2 + \cdots + R_n] = \mathrm{Ex}[R_1] + \mathrm{Ex}[R_2] + \cdots + \mathrm{Ex}[R_n].$$

Kỳ vọng của phân phối nhị thức

Câu hỏi

Tung n đồng xu, mỗi đồng có xác suất xảy ra mặt ngửa là p. Kỳ vọng của số mặt ngửa bằng bao nhiêu?

Đặt

$$J = \text{ số mặt ngửa}$$

Vậy thì J có phân phối nhị thức với tham số n và p:

$$\Pr[J = k] = \binom{n}{k} p^k (1-p)^{n-k}.$$

Một lời giải khó

$$\mathrm{Ex}[J] = \sum_{k=0}^{n} k \cdot \Pr[J = k] = \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k}$$
$$= pn.$$

Một lời giải dễ

Xét các biến ngẫu nhiên chỉ báo

$$J_i = \begin{cases} 1 & \text{nếu đồng xu thứ } i \text{ ngửa} \\ 0 & \text{ngược lại} \end{cases}$$

Vậy thì số mặt ngửa là

$$J = J_1 + J_2 + \cdots + J_n.$$

Ta được

$$\mathrm{Ex}[J\,] = \sum_{i=1}^{n} \Pr[J_i = 1] = pn.$$

Chứng minh bằng phương pháp xác suất

Ta cũng vừa chứng minh đẳng thức phức tạp

$$\sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k} = pn.$$

Bài toán (Coupon Collector)

- Giả sử mỗi hộp bỏng ngô có kèm một chiếc tem.
- Có n loại tem khác nhau: xanh, đỏ, tím, vàng, cam...
- Nếu bạn sưu tầm được đủ mỗi loại tem ít nhất một chiếc thì bạn sẽ được nhận phần thưởng.
- Giả sử chiếc tem trong mỗi hộp bỏng ngô được chọn ngẫu nhiên và theo phân phối đều.
- Hỏi rằng bạn phải mua bao nhiêu hộp bỏng ngô thi mới được mỗi loại ít nhất một chiếc tem?

Ý tưởng

- Giả sử ta có 5 loại tem và ta có dãy tem sau đây

  xanh đỏ đỏ tím xanh đỏ vàng xanh nâu

- Ta phân hoạch thành các đoạn

$$\underbrace{\text{xanh}}_{X_0}\quad\underbrace{\text{đỏ}}_{X_1}\quad\underbrace{\text{đỏ}\quad\text{tím}}_{X_2}\quad\underbrace{\text{xanh}\quad\text{đỏ}\quad\text{vàng}}_{X_3}\quad\underbrace{\text{xanh}\quad\text{nâu}}_{X_4}$$

  theo quy tắc kết thúc một đoạn ngay khi ta gặp loại tem mới.

Lời giải bài toán Coupon Collector

- Xét X_k là độ dài của đoạn thứ k.
- Tổng số hộp bỏng ngô ta phải mua để có đủ n loại tem là

$$T = X_0 + X_1 + X_2 + \cdots + X_{n-1}.$$

Kỳ vọng của X_k

- Bắt đầu đoạn thứ k ta có k loại tem khác nhau, và đoạn thứ k kết thúc khi ta có loại tem mới.
- Mỗi hộp bỏng ngô có thể kèm một kiểu tem mới với xác suất là
$$\frac{n-k}{n}.$$
- Vậy thì
$$\mathrm{Ex}[X_k] = \frac{n}{n-k}.$$

  Tại sao?

Lời giải bài toán Coupon Collector

$$\begin{aligned}
\mathrm{Ex}[T] &= \mathrm{Ex}[X_0 + X_1 + \cdots + X_{n-1}] \\
&= \mathrm{Ex}[X_0] + \mathrm{Ex}[X_1] + \cdots + \mathrm{Ex}[X_{n-1}] \\
&= \frac{n}{n-0} + \frac{n}{n-1} + \cdots + \frac{n}{3} + \frac{n}{2} + \frac{n}{1} \\
&= n\left(\frac{1}{n-0} + \frac{1}{n-1} + \cdots + \frac{1}{3} + \frac{1}{2} + \frac{1}{1}\right) \\
&\approx n \ln n.
\end{aligned}$$

Ví dụ

Trung bình ta cần tung con xúc xắc khoảng

$$6 \times (1 + 1/2 + 1/3 + 1/4 + 1/5 + 1/6) \approx 14.7$$

lần để có đủ cả 6 mặt xuất hiện.

Định lý

Xét không gian xác suất $\mathcal{S}$ và các sự kiện $A_1, A_2, \ldots, A_n \subseteq \mathcal{S}$, vậy kỳ vọng của số lượng sự kiện xuất hiện bằng

$$\Pr[A_1] + \Pr[A_2] + \cdots + \Pr[A_n]$$

Chứng minh.

Đặt

$$T_i(w) = \begin{cases} 1 & \text{nếu } w \in A_i \\ 0 & \text{ngược lại} \end{cases}$$

tức $T_i = 1$ nếu và chỉ nếu A_i xảy ra; và đặt

$$T = T_1 + T_2 + \cdots + T_n.$$

Ta có

$$\begin{aligned} \mathrm{Ex}[T] &= \sum_{i=1}^{n} \mathrm{Ex}[T_i] \\ &= \sum_{i=1}^{n} \Pr[T_i = 1] \\ &= \sum_{i=1}^{n} \Pr[A_i]. \end{aligned}$$

Ví dụ

Tung n đồng xu và đặt

$$\begin{aligned} A_i &= \text{ sự kiện đồng xu thứ } i \text{ ngửa} \\ T &= \text{ số mặt ngửa khi tung } n \text{ đồng} \end{aligned}$$

Ta có

$$\begin{aligned} \mathrm{Ex}[T] &= \Pr[A_1] + \Pr[A_2] + \cdots + \Pr[A_n] \\ &= 1/2 + 1/2 + \cdots + 1/2 \\ &= n/2 \end{aligned}$$

Chứng minh khó hơn

$$\begin{aligned}
\mathrm{Ex}[T] &= \sum_{i=1}^{n} i \cdot \Pr[T = i] \\
&= \sum_{i=1}^{n} i \cdot \binom{n}{i} 2^{-n} \qquad \text{(Giả sử các sự kiện này độc lập)} \\
&= n/2
\end{aligned}$$

Chứng minh bằng phương pháp xác suất

Như vậy ta đã chứng minh đẳng thức

$$\sum_{i=1}^{n} i\binom{n}{i} = n \cdot 2^{n-1}$$

Định lý

$$\underbrace{\Pr[T \geq 1]}_{\text{xác suất ít nhất một sự kiện } A_i \text{ xuất hiện}} \leq \mathrm{Ex}[T]$$

Chứng minh.

$$\begin{aligned}\mathrm{Ex}[T] &= \sum_{i=1}^{\infty} \Pr[T \geq 1] \\ &\geq \Pr[T \geq 1].\end{aligned}$$

□

Hệ quả

$$\Pr[T \geq 1] \leq \sum_{i=1}^{\infty} \Pr[A_i].$$

Ví dụ

- Nếu $n = 1000$ và $\Pr[A_i] = 1/100$, vậy thì
$$\mathrm{Ex}[T] = 10.$$
- Nếu với mọi i, j ta có $\Pr[A_i \mid A_j] = 1$, vậy thì
$$\Pr[T \geq 1] = 1/100 < 10.$$

Định lý (Luật Murphy)

Xét các sự kiện độc lập $A_1, \ldots, A_n$. Vậy thì

$$\Pr[T = 0] \le e^{-\mathrm{Ex}[T]}.$$

Chứng minh.

$$\begin{aligned}
\Pr[T = 0] &= \Pr[\overline{A_1} \cap \overline{A_2} \cap \cdots \cap \overline{A_n}] \\
&= \prod_{i=1}^{n} \Pr[\overline{A_i}] && \text{(do độc lập)} \\
&= \prod_{i=1}^{n} (1 - \Pr[A_i]) && \text{(do } 1 - x \le e^{-x}\text{)} \\
&\le \prod_{i=1}^{n} e^{-\Pr[A_i]} = e^{-\sum_{i=1}^{n} \Pr[A_i]} = e^{-\mathrm{Ex}[T]}.
\end{aligned}$$

□

Hệ quả

*Nếu mong đợi ít nhất 10 sự kiện độc lập xuất hiện, vậy thì xác suất **không** có sự kiện nào xuất hiện $\leq e^{-10} < 1/22,000$.*

Định lý (Luật tích)

Với mọi cặp biến ngẫu nhiên **độc lập** R_1, R_2

$$\mathrm{Ex}[R_1 \cdot R_2] = \mathrm{Ex}[R_1] \cdot \mathrm{Ex}[R_2].$$

Ví dụ

Tung hai con xúc xắc sáu mặt độc lập và đặt

$$R_i = \text{ kết quả của con xúc xắc } i$$

Ta có

$$\begin{aligned} \mathrm{Ex}[R_1 \cdot R_2] &= \mathrm{Ex}[R_1] \cdot \mathrm{Ex}[R_2] \\ &= 7/2 \cdot 7/2 = 49/4. \end{aligned}$$

Điều kiện độc lập là cần thiết

Ví dụ

$$
\begin{aligned}
\mathrm{Ex}[R_1 \cdot R_1] &= \mathrm{Ex}[R_1^2] \\
&= \sum_{i=1}^{6} i^2 \Pr[R_1 = i] \\
&= \frac{1}{16}(1 + 4 + 9 + 16 + 25 + 36) \\
&= 15\frac{1}{6} \neq \left(3\frac{1}{2}\right)^2 = \mathrm{Ex}[R_1]^2
\end{aligned}
$$

Hệ quả

Nếu $R_1, R_2, \ldots, R_n$ *là các biến ngẫu nhiên độc lập, vậy thì*

$$\mathrm{Ex}[R_1 \cdot R_2 \cdots R_n] = \mathrm{Ex}[R_1] \cdot \mathrm{Ex}[R_2] \cdots \mathrm{Ex}[R_n].$$

Chứng minh rằng: Với mọi hằng số a, b và mọi biến ngẫu nhiên R,

$$\mathrm{Ex}[a \cdot R + b] = a \cdot \mathrm{Ex}[R] + b.$$

Câu hỏi
Liệu đẳng thức

$$\mathrm{Ex}\left[\frac{1}{R}\right] = \frac{1}{\mathrm{Ex}[R]}$$

có đúng?

Nghịch lý RISC về kích thước mã nguồn

Benchmark	RISC	CISC	CISC/RISC
E-string search	150	120	0.8
F-bit test	120	180	1.5
Ackerman	150	300	2.0
Rec Sort	2800	1400	0.5
	3220	2000	

$$\text{Trung bình } [\ \text{CISC/RISC}\ ] = \frac{1}{4} \times (0.8 + 1.5 + 2.0 + 0.5)$$
$$= 1.2$$

Kết luận

"Các chương trình CISC là dài hơn 20% về trung bình."

Nghịch lý RISC về kích thước mã nguồn 2

Benchmark	RISC	CISC	RISC/CISC
E-string search	150	120	1.25
F-bit test	120	180	0.67
Ackerman	150	300	0.5
Rec Sort	2800	1400	2.0
Trung bình			1.1

Kết luận

"Các chương trình RISC là dài hơn 10% về trung bình."

Tại sao?

- Xét biến ngẫu nhiên R là độ dài của chương trình viết bằng RISC,
- và biến ngẫu nhiên C là độ dài của chương trình viết bằng CISC.
- Tính toán

$$\mathrm{Ex}[C/R] = 1.2 \qquad \checkmark$$

- Kết luận "chương trình CISC là dài hơn 20% về trung bình" có nghĩa rằng

$$\mathrm{Ex}[C] = 1.2\ \mathrm{Ex}[R] \qquad ✗$$

Một kết luận hợp lý

$$\begin{aligned}\mathrm{Ex}[R] &= \sum_{i \in Range(R)} i \cdot \Pr[R = i] \\ &= \frac{150}{4} + \frac{120}{4} + \frac{150}{4} + \frac{2800}{4} = 805.\end{aligned}$$

$$\begin{aligned}\mathrm{Ex}[C] &= \sum_{i \in Range(C)} i \cdot \Pr[C = i] \\ &= \frac{120}{4} + \frac{180}{4} + \frac{300}{4} + \frac{1400}{4} = 500.\end{aligned}$$

Vậy $\mathrm{Ex}[R]/\mathrm{Ex}[C] = 1.61$.

Kết luận

"Chương trình RISC là dài hơn 61% về trung bình."