C H Ư Ơ N G **4**

CÁC PHÂN PHỐI XÁC SUẤT RỜI RẠC HỮU DỤNG

***Về chương này*:**

Trong chương này, chúng tôi giới thiệu một số phân phối xác suất rời rạc được xem như là các mô hình cho những sự đo lường được thực hiện trong các điều kiện quan sát hay thí nghiệm đã phát sinh trong những lĩnh vực tiếp thị, kinh tế học hay kinh doanh nói chung. Điểm trọng tâm của chương này là sự ứng dụng của những phân phối này như là các mô hìnhtrong kinh doanh và kinh tế học.

NGHIÊN CỨU ĐIỂN HÌNH

LÁI XE - MỘT QUYỀN LỢI HAY MỘT ĐẶC QUYỀN

Tình yêu vĩnh cửu của người Mỹ dành cho xe cộ đã bén rễ vào trong toàn bộ đời sống của chúng ta. Số ngày này hầu như chẳng là bao nếu như có bất cứ một ngày nào mà một người Mỹ sở hữu xe hơi không ngồi vào đằng sau tay lái để lái xe đi làm, dùng chung xe để đưa rước bọn trẻ đi đến và đi về từ các hoạt động hàng ngày, chạy việc vặt, mua sắm, hay đơn giản là lái xe chỉ vì yêu thích. Tuy nhiên, theo Frank Newport và Leslie McAneny (1993) khi điều tra 1.003 người lớn vào tháng Sáu và 803 thiếu niên vào tháng Chín năm 1993, thì người Mỹ, cả người lớn lẫn thiếu niên, đều cảm thấy rằng bằng lái xe không phải là một quyền lợi mà là một đặc quyền. Họ tìm thấy rằng 70% số người lớn được hỏi ủng hộ một kỳ thi mang tính bắt buộc mỗi ba năm một đối với những người trên 65 tuổi và rằng 56% số thiếu niên được điều tra đã ủng hộ cho các điều luật ở tiểu bang mà sẽ từ chối cấp bằng lái xe cho những ai dưới 21 tuổi mà đã bỏ học trung học. Báo cáo của hai tác giả này khẳng định rằng các tỷ lệ phần trăm được ghi nhận đối với người lớn chỉ khác với tỷ lệ phần trăm thực tế đối với toàn bộ số người lớn một mức không lớn hơn 3 điểm phần trăm và rằng các tỷ lệ phần trăm được báo cáo cho thiếu niên chỉ khác với tỷ lệ phần trăm thực tế của toàn bộ số thiếu niên không nhiều hơn 4 điểm phần trăm.

Bằng cách nào mà chúng ta có thể chắc chắn rằng các tỷ lệ phần trăm được báo cáo là chính xác giống như điều được khẳng định. Khi các cuộc điều tra được thực hiện bằng cách sử dụng các câu trả lời có và không, thì các câu hỏi rõ ràng dành cho một sinh viên thống kê sẽ là "Mô hình thống kê nào là thích hợp trong những tình huống giống như thế này" và, thứ hai, "Bằng cách nào mà chúng ta có thể sử dụng các mô hình này để đánh giá độ tin cậy của những kết luận dựa trên các câu trả lời cho những câu hỏi có và không?"

Những câu hỏi này các các câu hỏi khác có liên quan đến các kết luận sẽ được đề cập trong các Phần 4.2 và 4.3 khi chúng tôi giới thiệu các phân phối nhị thức và phân phối Poisson. Hãy nhớ rằng các kết luận được căn cứ trên trung bình và độ lệch chuẩn, mà giá trị của chúng được xác định qua việc sử dụng thông tin chọn mẫu. Chúng tôi sẽ quay trở lại câu hỏi "Mức độ tin cậy cao đến đâu trong sự ước tính tỷ lệ phần trăm các tài xế mà ủng hộ cho những bài kiểm tra mang tính bắt buộc đối với các công dân đã trưởng thành?"

4.1 GIỚI THIỆU

Trong Chương 3, chúng tôi đã tìm thấy rằng các số hạng ngẫu nhiêu được xác định qua một con số hữu hạn hay một con số vô hạn có khả năng đếm được của các sự kiện đơn giản được gọi là **các biến số ngẫu nhiên rời rạc.** Những ví dụ của những biến số ngẫu nhiên rời rạc thì đầy dẫy trong kinh doanh và kinh tế học, nhưng chỉ có ba phân phối xác suất rời rạc đóng vai trò như là **các mô hình** cho một con số lớn các ứng dụng. Ba phân phối này là phân phối xác suất **nhị thức, Poisson, và siêu bội.** Trong chương này, chúng ta nghiên cứu những phân phối này và thảo luận sự tiến triển của chúng như là các mô hình hợp lý cho các qui trình rời rạc được quan sát thấy trong tự nhiên.

4.2 PHÂN PHỐI XÁC SUẤT NHỊ THỨC

Một trong những biến số ngẫu nhiên rời rạc cơ bản, hữu ích và thú vị nhất - biến số ngẫu nhiên nhị thức - được kết hợp với thí nghiệm tung đồng tiền như được mô tả trong các Ví dụ 3.2 và 3.12. Để minh họa, hãy xem xét một cuộc điều tra chọn mẫu được thực hiện nhằm xác định sự chấp nhận của thị trường đối với một sản phẩm mới. Mỗi người được phỏng vấn là tương tự như việc tung đồng tiền, bởi vì sự chấp nhận của một người đối với một sản phẩm là tương tự với việc quan sát một mặt ngửa, và rằng sự từ chối của người đó đối với sản phẩm đó là tương tự như việc quan sát một mặt sấp. Sự khác biệt ở đây là xác suất của sự chấp nhận một sản phẩm mới thường không phải là 1/2.

Các cuộc điều tra tương tự được thực hiện trong các ngành khoa học xã hội, công nghiệp và tiếp thị. Nhà nghiên cứu xã hội học quan tâm đến tỷ lệ các công dân gốc Tây Ban Nha và Bồ Đào Nha có đăng ký bầu cử; nhà sản xuất các bo mạch in thì quan tâm đến số bo mạch có ít nhất một khuyết điểm; một nhóm môi trường thì quan tâm đến tỷ lệ phần trăm các gia đình tham gia tích cực vào việc tái chế các lon nhôm. Mặc dù có sự khác biệt về một số phương diện, thì các thí nghiệm được mô tả ở đây thường thể hiện, ở một mức độ tương tự hợp lý, các đặc trưng của một **thí nghiệm nhị thức.**

ĐỊNH NGHĨA Một **thí nghiệm nhị thức** là thí nghiệm mà có các đặc trưng sau:

1. Thí nghiệm đó có *n* lần thử giống nhau.
2. Mỗi lần thử tạo ra một trong hai kết quả. Do không có thuật ngữ nào tốt hơn, chúng ta sẽ gọi kết quả thành công là *S* và kết quả thất bại là *F*.[+]
3. Xác suất của thành công trong một lần thử duy nhất là bằng với *p* và giữ nguyên như vậy qua các lần thử nghiệm. Xác suất của thất bại là bằng với (1-*p*) = *q*.
4. Các lần thử là độc lập với nhau.
5. Chúng ta quan tâm đến *x*, con số những lần thành công được quan sát trong *n* lần thử.

VÍ DỤ **4.1** Giả định rằng có khoảng 1 triệu người lớn trong một khu vực bán hàng nào đó mà là người mua tiềm năng của một sản phẩm mới và rằng một tỷ phần chưa biết *p* sẽ mua sản phẩm này nếu như nó được đưa ra chào bán. Một mẫu gồm 1.000 người lớn sẽ được chọn theo một cách thức mà mỗi người trong số 1 triệu người trong khu vực bán hàng này sẽ có một cơ hội ngang nhau của việc được chọn lựa. Mỗi người lớn trong mẫu này sẽ được hỏi rằng liệu ông/bà ta sẽ mua sản phẩm này nếu như nó được đưa ra chào bán. (Mục đích cuối cùng của cuộc điều tra này là ước lượng tỷ phần chưa biết *p*, một bài toán mà chúng ta sẽ học cách xử lý trong Chương 7). Liệu đây có phải là một thí nghiệm nhị thức?

Lời giải Để quyết định liệu đây có phải là một thí nghiệm nhị thức hay không, chúng ta phải xem xét rằng liệu việc chọn mẫu có thỏa mãn năm đặc trưng được mô tả trong phần định nghĩa ở trên.

1. Việc chọn mẫu này bao gồm *n* = 1.000 lần thử giống nhau. Một lần thử tượng trưng cho sự chọn lựa một người lớn duy nhất từ 1 triệu người lớn trong khu vực bán hàng.
2. Mỗi lần thử sẽ tạo ra một trong hai kết quả. Một người sẽ khẳng định rằng hoặc ông/bà ta sẽ mua sản phẩm mới này hoặc ông/bà ta không mua. Hai kết quả này có thể được liên tưởng đến sự "thành công" hay "thất bại" của một thí nghiệm nhị thức.
3. Xác suất của một sự thành công sẽ bằng với tỷ lệ của 1 triệu người lớn mà sẽ mua sản phẩm mới này. Ví dụ, nếu 500.000 trong số 1 triệu người lớn trong khu vực này sẽ mua sản phẩm đó, thì xác suất mà người lớn đầu tiên được chọn sẽ mua sản phẩm này là *p* = 0,5. Trên thực tế, xác suất này sẽ giữ nguyên từ lần thử này sang lần thử khác, thậm chí khi những người lớn được chọn lựa trong các lần thử trước đó không bị thay thế khi việc chọn mẫu vẫn tiếp diễn.

[+] Mặc dù người ta thường gọi hai kết quả có thể có của một lần thử là "thành công" và "thất bại", thì chúng cũng có thể được gọi là "ngửa" và "sấp", "đỏ" và "trắng", hay bất cứ một cặp từ ngữ nào khác. Do đó, kết quả được gọi là thành công không nhất thiết được xem như là một sự thành công trong nghĩa thông thường của từ này.

4. Trên thực tế, xác suất của sự thành công trong bất cứ lần thử nào sẽ không bị tác động bởi kết quả của bất kỳ lần thử khác (nó sẽ rất gần với *p*).
5. Chúng ta quan tâm đến con số *x* của người lớn trong mẫu của 1.000 người mà sẽ mua sản phẩm này.

Bởi vì cuộc điều tra này thỏa mãn tương đối tốt năm đặc trưng trên, cho nên trên thực tế thì cuộc điều tra này (cũng giống như nhiều cuộc trưng cầu ý kiến khác) có thể được xem như một thí nghiệm nhị thức.

VÍ DỤ **4.2** Một người mua hàng, người đã nhận được chuyến hàng gởi bao gồm 20 máy tính cá nhân, mong muốn chọn mẫu 3 trong số các máy tính này để xem liệu chúng có hoạt động tốt hay không trước khi dỡ hàng. Ba máy tính gần nhất được đem ra chạy thử, và sau đó, được tuyên bố là hoặc mắc lỗi hoặc không có khiếm khuyết gì. Điều mà người mua hàng không biết là, 2 trong số 20 máy tính này là có lỗi. Liệu đây có phải là một thí nghiệm nhị thức?

Lời giải Giống như trong Ví dụ 4.1, chúng ta kiểm tra thủ tục chọn mẫu so với các đặc trưng của một thí nghiệm nhị thức.

1. Thí nghiệm này bao gồm $n = 3$ lần thử giống nhau. Mỗi lần thử tượng trưng cho sự chọn lựa và kiểm tra của một máy tính trong tổng số 20 máy tính.
2. Mỗi lần thử tạo ra một trong hai kết quả. Hoặc máy tính đó là có lỗi (gọi đây là một "thành công") hoặc không phải (một "thất bại").
3. Giả định rằng các máy tính được chất ngẫu nhiên lên một côngtenơ chở hàng, để cho bất cứ máy tính nào trong số 20 máy tính này ắt có thể đã được đặt gần cửa côngtenơ. Sau đó xác suất vô điều kiện của việc chọn ra một máy tính có lỗi trên một lần thử sẽ là 2/20.
4. Điều kiện của sự độc lập giữa các lần thử *không* được thỏa mãn bởi vì xác suất của việc chọn ra một máy tính có lỗi vào các lần thử thứ hai và thứ ba sẽ phụ thuộc vào kết quả của lần thử đầu tiên. Ví dụ, nếu lần thử đầu tiên cho ra kết quả là một máy tính có lỗi, thì sau đó chỉ còn lại một máy tính có lỗi trong số 19 máy tính còn lại trong lô hàng. Vì thế, xác suất có điều kiện của thành công ở lần thử 2, khi đã biết sự thành công trong lần thử 1, sẽ là 1/19. Kết quả này khác với xác suất vô điều kiện cho một sự thành công trong lần thử thứ hai (mà là 2/20). Như vậy, các lần thử là phụ thuộc nhau và việc chọn mẫu này không tượng trưng cho một thí nghiệm nhị thức.

Ví dụ 4.2 minh họa một điểm quan trọng. Nếu cỡ mẫu *n* là lớn so với qui mô tổng thể *N*, thì xác suất của sự thành công *p* sẽ không còn giữ nguyên qua các lần thử nữa. Như vậy, các kết quả của những lần thử sẽ phụ thuộc nhau, và thí nghiệm tạo ra sẽ không phải là một thí nghiệm nhị thức. **Theo kinh nghiệm thực tế, nếu $n/N \geq 0,05$; thì thí nghiệm tạo ra sẽ không phải là nhị thức.**

Phân phối xác suất của một biến số ngẫu nghiên nhị thức đơn giản (số lượng "mặt ngửa" trong những lần tung của hai đồng tiền) được suy ra từ Ví dụ 3.12. Phân phối xác suất cho một thí nghiệm nhị thức bao gồm *n* số lần tung được suy ra chính xác theo cùng cách này, nhưng việc này là phức tạp hơn nhiều khi số *n* của các lần thử là lớn. Chúng ta sẽ bỏ qua sự suy ra này và chỉ đơn giản trình bày **phân phối xác suất nhị thức** và số trung bình, phương sai và độ lệch chuẩn của nó, như được trình bày trong mô tả dưới đây.

Phân phối Xác suất Nhị thức

$$p(x) = C_x^n p^x q^{n-x} = \frac{n!}{x!(n-x)!} p^x q^{n-x}$$

x, con số thành công trong n lần thử, có thể có các giá trị 0, 1, 2,..., n; p là xác suất của thành công trong một lần thử duy nhất; và C_x^n được xác định bằng:

$$C_x^n = \frac{n!}{x!(n-x)!}$$

trong đó $n! = n(n-1)(n-2)...(2)(1),$ và $0! \equiv 1$.

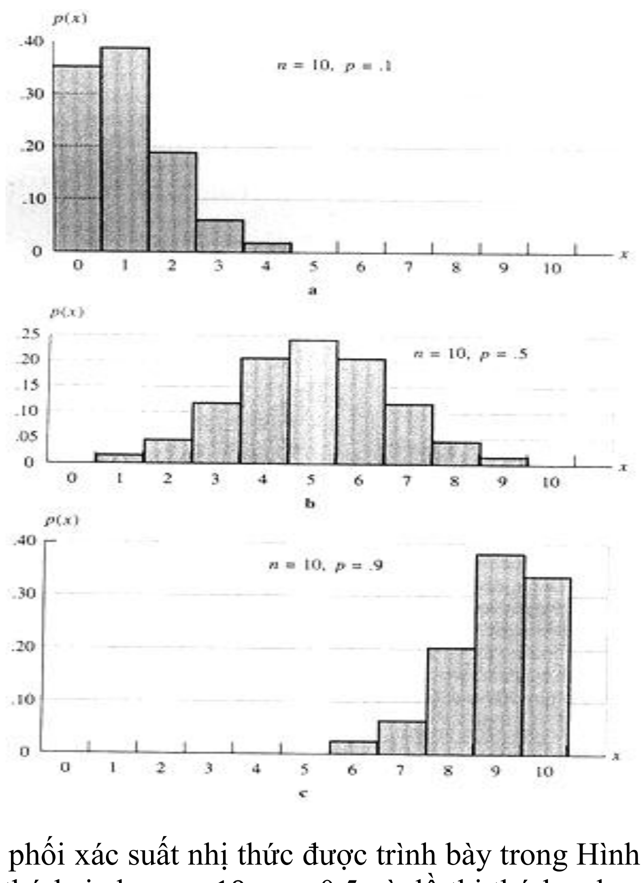
Trung bình: $\mu = np$

Phương sai: $\sigma^2 = npq$

Độ lệch chuẩn: $\sigma = \sqrt{npq}$

Ký hiệu giai thừa n! được sử dụng để đại diện cho tích số n(n-1) (n-2)... (3)(2)(1). Ví dụ, 5! = (5)(4)(3)(2)(1) = 120 và 0! được định nghĩa là bằng với 1. Ký hiệu C_x^n **là viết tắt cho *n*!/[*x*!(*n*-*x*)!],** một biểu thức mà xuất hiện trong công thức cho sự phân phối xác suất nhị thức.

HÌNH 4.1 **Các phân phối xác suất nhị thức**



Đồ thị của ba phân phối xác suất nhị thức được trình bày trong Hình 4.1: đồ thị đầu tiên cho $n = 10, p = 0{,}1$; đồ thị thứ hai cho $n = 10, p = 0{,}5$ và đồ thị thứ ba cho $n = 10, p = 0{,}9$. Độ cao của thanh đứng cho bất cứ giá trị nào của x được tính toán bằng cách sử dụng công thức nhị thức, $p(x)$ được cho ở trên. Lưu ý rằng khi $p = 0{,}5$ thì phân phối này là đối xứng; nếu p nhỏ, thì phân phối này bị lệch về bên phải; nếu p lớn thì phân phối này bị lệch về bên trái.

VÍ DỤ **4.3** Các nghiên cứu đã cho thấy rằng cứ một trong mỗi năm người sống ở căn hộ đã di chuyển chỗ ở trong vòng một năm cho trước. Giả định rằng có bốn cư dân căn hộ được phỏng vấn.

a. Xác suất mà chính xác là hai người đã chuyển chỗ ở trong vòng một năm qua là bao nhiêu?

b. Xác suất mà có ít nhất hai người đã chuyển chỗ ở trong vòng một năm qua là bao nhiêu?

c. Xác suất mà tất cả bốn người đã chuyển chỗ ở trong vòng một năm qua là bao nhiêu?

Lời giải

a. Định nghĩa biến số ngẫu nhiên *x* là con số của bốn người sống ở căn hộ mà đã di chuyển chỗ ở trong vòng một năm qua. Giả định rằng bốn người sống ở căn hộ này đã được chọn lựa một cách độc lập và không có liên hệ gì với nhau, thì *p* vẫn không thay đổi từ người này sang người khác và *x* là một biến số ngẫu nhiên nhị thức với $n = 4$ và $p = 2$. Vì vậy,

$$\begin{aligned}
p(x) &= C_x^4(0,2)^x(0,8)^{4-x} \\
p(2) &= C_2^4(0,2)^2(0,8)^{4-2} \\
&= \frac{4!}{2!2!}(0,04)(0,64) = \frac{4(3)(2)(1)}{2(1)(2)(1)}(0,04)(0,64) \\
&= 0,1536
\end{aligned}$$

b.

$$\begin{aligned}
P(toi\ thieu\ hai) &= p(2) + p(3) + p(4) \\
&= 1 - p(0) - p(1) \\
&= 1 - C_0^4(0,2)^0(0,8)^4 - C_1^4(0,2)(0,8)^3 \\
&= 1 - 0,4096 - 0,4096 \\
&= 0,1808
\end{aligned}$$

c.

$$\begin{aligned}
p(4) &= C_4^4(0,2)^4(0,8)^0 \\
&= \frac{4!}{4!0!}(0,2)^4(1) = 0,0016
\end{aligned}$$

VÍ DỤ **4.4** Những lô lớn các sản phẩm sắp ra mắt tại một nhà máy sản xuất được kiểm tra để tìm lỗi bằng cách sử dụng *phương án chọn mẫu*. Một mẫu ngẫu nhiên gồm *n* sản phẩm được chọn lựa từ mỗi lô hàng và được kiểm tra, và số *x* của các sản phẩm sai sót trong mẫu đó được ghi nhận. Nếu *x* là nhỏ hơn hay bằng với *một con số chấp nhận a* nào đó đã được xác định trước, thì lô hàng đó được chấp nhận. Nếu *x* lớn hơn *a*, thì lô hàng đó bị từ chối. Giả định rằng một nhà sản xuất sử dụng một phương án chọn mẫu với $n = 10$ và $a = 1$. Nếu một lô hàng chứa đựng chính xác 5% sản phẩm bị lỗi, thì xác suất để cho lô hàng đó được chấp thuận là bao nhiêu? từ chối là bao nhiêu?

Lời giải Bởi vì lô hàng đó chứa 5% sản phẩm bị lỗi, nên xác suất mà một sản phẩm được rút từ lô hàng đó mà bị lỗi là $p = 0,5$. Cho nên xác suất của việc quan sát thấy *x* sản phẩm bị lỗi trong một mẫu gồm $n = 10$ sản phẩm là:

$$p(x) = C_x^{10}(0.05)^x(0,95)^{10-x}$$

Xác suất của việc chấp thuận lô hàng này là xác suất rằng *x* nhỏ hơn hay bằng với con số được chấp nhận *a* = 1. Do đó,

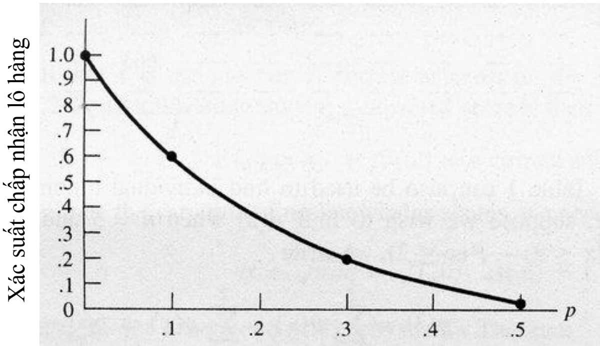
$$P(\text{chap nhan}) = p(0) + p(1) = C_0^{10}(0{,}05)^0(0{,}95)^{10} + C_1^{10}(0{,}05)^1(0{,}95)^9 = 0{,}914$$

$$P(\text{tu choi}) = 1 - P(\text{chap nhan}) = 1 - 0{,}914 = 0{,}086$$

Mặc dù trong tình huống thực tế chúng ta sẽ không biết giá trị chính xác của *p*, chúng ta cũng sẽ muốn biết xác suất của việc chấp thuận các lô hàng kém (các lô hàng mà có *p* lớn) và những lô hàng tốt (các lô hàng mà có *p* nhỏ). Ví dụ này cho thấy cách thức mà chúng ta sẽ tính toán xác suất chấp nhận này cho các giá trị khác nhau của *p*.

Một đồ thị của xác suất cho sự chấp nhận lô hàng so với tỷ lệ bị lỗi của lô hàng đó *p* được gọi là **đường cong đặc trưng hoạt động** cho phương án chấp thuận một lô hàng. Đường cong đặc trưng hoạt động cho một phương án chọn mẫu với *n* = 5 và *a* = 0 được trình bày trong Hình 4.2.

HÌNH 4.2 **Đường cong đặc trưng hoạt động cho *n* = 5 và *a* = 0**



Lưu ý rằng việc chọn mẫu chấp nhận là một ví dụ của sự kết luận về mặt thống kê bởi vì việc này hàm ý một quyết định liên quan đến tỷ lệ bị lỗi của lô hàng *p*. Nếu bạn chấp nhận một lô hàng, thì bạn kết luận rằng tỷ lệ thực tế bị lỗi của lô hàng *p* là tương đối nhỏ, có giá trị chấp nhận được. Nếu bạn từ chối lô hàng đó, rõ ràng là bạn nghĩ rằng *p* quá lớn. Do vậy, việc chọn mẫu chấp nhận lô hàng là một qui trình mà hàm ý việc suy ra sự kết luận có liên quan đến tỷ lệ bị lỗi của lô hàng. Đường cong đặc trưng hoạt động cho phương án chọn mẫu này cung cấp cho ta một thước đo về mức độ tốt của quá trình suy ra kết luận này.

Tính toán các xác suất nhị thức là một nhiệm vụ nặng nhọc khi *n* lớn. Để đơn giản hóa các tính toán của chúng ta, tổng các xác suất nhị thức từ *x* = 0 đến *x* = *a* được trình bày trong Bảng 1 của Phụ lục II đối với *n* = 2, 3,..., 12, 15, 20 và 25.

Để minh họa sự sử dụng của Bảng 1, lãy tìm ra tổng của các xác suất nhị thức từ *x* = 0 đến *x* = 3 cho *n* = 5 lần thử và *p* = 0,6. Nghĩa là, chúng ta mong muốn tìm ra:

$$P(x \le 3) = \sum_{x=0}^{3} p(x) = p(0) + p(1) + p(2) + p(3)$$

trong đó

$$p(x) = C_x^5 (0{,}6)^x (0{,}4)^{5-x}$$

Bởi vì các giá trị trong bảng cho chúng ta

$$P(x \le a) = \sum_{x=0}^{a} p(x)$$

cho nên chúng ta tìm kiếm giá trị bảng trong hàng tương ứng với a = 3 và cột cho p = 0,6. Giá trị bảng, 0,663 được trình bày trong Bảng 4.1 như đã xuất hiện trong Bảng 1, Phụ lục II. Như vậy, tổng của các xác suất nhị thức từ x = 0 đến x = a = 3 (đối với n = 5, p = 0,6).

BẢNG 4.1 **Phần của Bảng 1, Phụ lục II đối với *n* = 5**

a	0,01	0,05	0,10	0,20	0,30	0,40	0,50	0,60	0,70	0,80	0,9	0,95	0,99	a
0	____	____	____	____	____	____	____	____	____	____	____	____	____	0
1	____	____	____	____	____	____	____	____	____	____	____	____	____	1
2	____	____	____	____	____	____	____	____	____	____	____	____	____	2
3	____	____	____	____	____	____	____	0,663	____	____	____	____	____	3
4	____	____	____	____	____	____	____	____	____	____	____	____	____	4

Bảng 1 cũng có thể được sử dụng để tìm ra các xác suất nhị thức riêng lẻ. ví dụ, giả định rằng chúng ta mong muốn tìm ra $p(3)$ khi n = 5 và p = 0,6. Bởi vì $P(x = 3) = P(x \le 3) - P(x \le 2)$, chúng ta viết lại:

$$p(3) = \sum_{x=0}^{3} p(x) - \sum_{x=0}^{2} p(x) = 0{,}663 - 0{,}317 = 0{,}346$$

Giá trị $\sum_{x=0}^{3} p(x)$ và $\sum_{x=0}^{2} p(x)$ được tìm ra trực tiếp từ Bảng 1, qua việc xem n = 5 với p = 0,6. Nói chung, một xác suất nhị thức riêng lẻ có thể được tìm ra bằng cách trừ các số liên tiếp trong bảng cho một giá trị cho trước của p.

VÍ DỤ **4.5** Giả định rằng bạn là giám đốc nhân sự của một công ty và rằng bạn mong muốn đánh giá điểm số trong một bài trắc nghiệm chọn câu trả lời đúng trong số nhiều lựa chọn để kiểm tra khả năng của những người dự tuyển xin việc. Một điểm số zêrô đối với một bài kiểm tra mục tiêu (các câu hỏi đòi hỏi sự nhớ lại toàn bộ tài liệu) chỉ ra rằng người đó không có khả năng nhớ lại tài liệu bài kiểm tra ở thời điểm bài kiểm tra được phát ra. Trái lại, một người với ít hay không có khả năng nhớ lại về bài kiểm tra có thể ghi điểm số cao hơn trong một bài kiểm tra chọn câu đúng trong số nhiều lựa chọn bởi vì người náy chỉ cần nhận ra (trái với phải nhớ lại) câu trả lời đúng và bởi vì một số câu hỏi sẽ được trả lời một cách chính xác chỉ bằng cơ may, ngay cả khi người đó không biết câu trả lời đúng. Do vậy, điểm số do không có kiến thức đối với một bài kiểm tra chọn câu trả lời đúng trong số nhiều chọn lựa ắt sẽ cao hơn zêrô. Giả định rằng một bài kiểm tra chọn câu trả lời đúng trong số nhiều chọn lựa bao gồm 20 câu hỏi, mỗi câu có 5 chọn lựa trả lời có thể có. Xác suất mà một người không có kiến thức về tài liệu kiểm tra trả lời chính xác được tám câu hỏi trở lên là bao nhiêu?

Lời giải Giả định rằng người đó không có kiến thức về câu trả lời đối với một câu hỏi, thì xác suất p rằng anh/chị ta trả lời chính xác câu hỏi đó là p = 0,2. Phân phối xác suất cho x, số lượng các câu trả lời chính xác cho một bài kiểm tra gồm 20 câu hỏi, là:

$$p(x) = C_x^{20}(0{,}2)^x(0{,}8)^{20-x}$$

Bởi vì việc tính toán trực tiếp khi n = 20 là rất mất thời gian, thay vào đó chúng ta chọn việc sử dụng Bảng 1, với n = 20 và p = 0,2. Sau đó:

$$P(x \ge 8) = 1 - P(x \le 7) = 1 - 0{,}968 = 0{,}032$$

trong đó $P(x \le 7)$ được tìm thấy trong Bảng 1 bằng cách tra số cho n = *20*, p = 0,2 và a = 7.

VÍ DỤ **4.6** Tham chiếu lại Ví dụ 4.5. Đâu là điểm số kỳ vọng cho một người mà không có kiến thức về tài liệu kiểm tra? Bạn sẽ kỳ vọng một điểm số không có kiến thức nằm trong những giới hạn nào?

Lời giải Nhắc lại rằng x là con số các câu trả lời chính xác trong một bài kiểm tra gồm 20 câu hỏi, với p = 0,2. Một điểm số kỳ vọng của ứng viên không có kiến thức sẽ là:

$$E(x) = np = 20\,(0{,}20) = 4 \text{ câu trả lời chính xác}$$

Để đánh giá sự biến thiên trong các điểm số không có kiến thức, chúng ta cần biết σ, trong đó

$$\sigma = \sqrt{npq} = \sqrt{(20)(0{,}2)(0{,}8)} = 1{,}79$$

Sau đó, từ kiến thức về Định lý Tchebysheff [+], chúng ta sẽ kỳ vọng rằng x rơi vào khoảng ($\mu \pm 2\sigma$) với một xác suất ít nhất là 0,75, và bên trong khoảng ($\mu \pm 3\sigma$) với xác suất tối thiểu 0,89. Những khoảng này là:

$$\mu \pm 2\sigma = 4 \pm 3{,}58 \text{ hay } 0{,}42 \text{ đến } 7{,}58$$

$$\mu \pm 3\sigma = 4 \pm 5{,}37 \text{ hay } -1{,}37 \text{ đến } 9{,}37$$

Con số này so với điểm số zêrô của ứng viên không có kiến thức khi làm bài kiểm tra nhớ lại.

Bạn cũng có thể sử dụng Minitab hay Excel để tạo ra các xác suất nhị thức riêng lẻ, $P(x = K)$, hay các xác suất nhị thức cộng dồn $P(x \leq K)$. (Chữ K đóng vai trò của chữ a trong Bảng 1 của Phụ lục II). Trong Minitab, sử dụng **Calc → Probability Distributions → Binomial,** và chọn hoặc "Probability" hoặc "Cumulative Probability". Bạn cũng có thể hoặc nhập vào một giá trị duy nhất cho K trong hộp có ghi chú "Input constant"; hoặc bạn có thể đánh vào một loạt các giá trị của x từ 0 đến n vào một cột trong cửa sổ Data, và nhập tên cột vào trong hộp có ghi chú "Input column". Nhập các giá trị cho n và p vào trong các hộp phù hợp, và bấm **OK.**

Trong Excel, sử dụng **Insert → Function → Statistical → BINOMDIST,** chọn dãy các ô có chứa các giá trị x, giá trị cho n và p, và chọn lựa của bạn cho các xác suất cộng dồn hay riêng lẻ. Để biết một sự giải thích chi tiết hơn, tham khảo phần **Sử dụng Excel** trong Phụ lục V.

Kết quả Minitab trong Bảng 4.2 cho thấy rằng cả xác suất nhị thức cộng dồn lẫn riêng lẻ đối với $n = 20$ và $p = 0{,}2$. Lưu ý rằng, trong kết quả Minitab, $P(x \leq 7) = 0{,}9679$ để cho $P(x \geq 8) = 1 - 0{,}9679 = 0{,}321$ mà, ở độ chính xác ba con số thập phân, thì khớp với các kết quả trước đó của chúng ta bằng cách sử dụng Bảng 1.

BẢNG 4.2. **Kết quả Minitab của các xác suất nhị thức khi *n = 20* và *p* = 0,2**

NHỊ THỨC VỚI N = 20 F = 20000		NHỊ THỨC VỚI N = 20 F = 20000	
K	P (X = K)	K	P (X ≤ K)
0	0,0115	0	0,0115
1	0,0576	1	0,0692
2	0,1369	2	0,2061
3	0,2054	3	0,4114
4	0,2182	4	0,6296
5	0,1746	5	0,8042
6	0,1091	6	0,9133
7	0,0545	7	0,9679
8	0,0222	8	0,9900
9	0,0074	9	0,9974

[+] Một biểu đồ tần suất của $p(x)$ đối với $n = 20$, $p = 0{,}2$ sẽ có hình dạng gò. Như thế, chúng ta sẽ kỳ vọng rằng Qui tắcThực chứng cũng sẽ vận hành rất tốt. Lý do tại sao như vậy sẽ được giải thích trong Chương 6.

10	0,0020	10	0,9994
11	0,0005	11	0,9999
12	0,0001	12	1,0000
13	0,0000		

Lưu ý rằng chúng ta đã lựa chọn không liệt kê các xác suất cho tất cả giá trị của x = 0, 1, 2,..., n bởi vì các xác suất cho tất cả giá trị của x từ 13 đến 20 là bằng với zêrô, khi làm tròn đến độ chính xác bốn con số thập phân. Đối với tất cả các giá trị này, xác suất nhị thức cộng dồn, $P(x \le K)$ sẽ bằng với 1,0000.

Bài tập

Các Kỹ thuật Cơ bản

4.1 Một lọ chứa 5 quả bóng - ba quả màu đỏ và hai quả màu trắng. Hai quả được chọn ngẫu nhiên mà không có sự thay thế từ cái lọ và con số x của các quả bóng màu đỏ được ghi nhận. Giải thích tại sao x là và không phải là một biến số ngẫu nhiên nhị thức. [*Gợi ý:* So sánh các đặc trưng của thí nghiệm này với những đặc thù của một thí nghiệm nhị thức được trình bày trong Phần 4.2]. Nếu thí nghiệm này là nhị thức, hãy cho biết các giá trị của n và p.

4.2 Trả lời Bài tập 4.1 bằng cách giả định rằng việc chọn mẫu được thực hiện với sự thay thế. Nghĩa là, quả bóng đầu tiên được chọn từ cái lọ và được quan sát. Sau đó nó được thay thế, và các quả bóng được trộn lẫn với nhau trước khi quả bóng thứ hai được chọn.

4.3 Tìm các kết quả cho:

a 3! **b** 5! **c** C_4^6

d C_3^8 **e** C_2^{10} **f** C_6^6

g C_0^6 **h** C_1^9 **i** C_0^9

4.4 Tính toán giá trị của $p(x)$ và xây dựng một biểu đồ xác suất cho:

a $n = 5, p = 0{,}2$ **b** $n = 5, p = 0{,}5$ **c** $n = 5, p = 0{,}8$

4.5 Tính toán $p(x)$ cho x =0, 1, 2,...5, 6 cho n = 6 và p = 0,1. Vẽ đồ thị $p(x)$. Lặp lại những chỉ dẫn này cho các phân phối xác suất nhị thức với p = 0,5 và p = 0,9. So sánh các đồ thị này. Bằng cách nào mà giá trị của p tác động đến hình dạng của $p(x)$?

4.6 Sử dụng Bảng 1 trong Phụ lục II để tìm ra tổng các xác suất nhị thức từ x = 0 đến x = a cho:

a $n = 8, p = 0{,}1, a = 3$ **b** $n = 12, p = 0{,}6, a = 7$ **c** $n = 25, p = 0{,}5, a = 14$

4.7 Sử dụng công thức cho phân phối xác suất nhị thức để tính toán các giá trị của $p(x)$ cho n = 5, p = 0,5 (sự tính toán này được thực hiện trong Bài tập 4.4). Sau đó tìm $\sum_{x=0}^{a} p(x)$ cho a = 0,1,2,3,4 bằng cách sử dụng các giá trị của $p(x)$ mà bạn đã tính ra. Kế đến so sánh các tổng này với những giá trị đã cho trong Bảng 1 trong Phụ lục II.

4.8 Sử dụng thông tin đã biết trong Bảng 1 trong Phụ lục II để tìm ra $p(3)$ cho n = 5, p = 0,5. Sau đó so sánh câu trả lời này với giá trị của $p(3)$ được tính trong Bài tập 4.7.

4.9 Sử dụng thông tin đã biết trong Bảng 1 trong Phụ lục II để tìm ra $p(3)$ + p(4) cho $n = 5$, $p = 0{,}5$. Kiểm tra câu trả lời này với các giá trị của $p(x)$ mà bạn đã tính trong Bài tập 4.7.

4.10 Sử dụng Bảng 1 trong Phụ lục II để tìm kết quả cho:

a $P(x < 12)$ cho $n = 20$, $p = 0{,}5$ **b** $P(x \le 6)$ cho $n = 15$, $p = 0{,}4$

c $P(x > 4)$ cho $n = 10$, $p = 0{,}4$ **d** $P(x \ge 6)$ cho $n = 20$, $p = 0{,}6$

e $P(3 < x < 7)$ cho $n = 10$, $p = 0{,}5$

4.11 Sử dụng Bảng 1 trong Phụ lục II để tìm kết quả cho:

a $P(x \le 5)$ cho $n = 10$, $p = 0{,}4$ **b** $P(x < 3)$ cho $n = 5$, $p = 0{,}6$

c $P(x \le 17)$ cho $n = 20$, $p = 0{,}7$ **d** $P(x > 17)$ cho $n = 20$, $p = 0{,}7$

e $P(x < 6)$ cho $n = 15$, $p = 0{,}4$

4.12 Tìm trung bình và độ lệch chuẩn cho các phân phối nhị thức sau đây:

a $n = 1000$, $p = 0{,}3$ **b** $n = 400$, $p = 0{,}01$

c $n = 500$, $p = 0{,}5$ **d** $n = 1600$, $p = 0{,}8$

4.13 Tìm trung bình và độ lệch chuẩn cho một phân phối nhị thức với $n = 100$ và các giá trị sau của p.

a $p = 0{,}01$ **b** $p = 0{,}9$ **c** $p = 0{,}3$

d $p = 0{,}7$ **e** $p = 0{,}5$

4.14 Trong Bài tập 4.13 chúng ta đã tính toán trung bình và độ lệch chuẩn cho một biến số ngẫu nhiên nhị thức đối với một cỡ mẫu cố định, $n = 100$, và cho các giá trị khác nhau của p. Hãy vẽ đồ thị cho các giá trị của độ lệch chuẩn đối với năm giá trị của p đã biết trong Bài tập 4.13. Với giá trị nào của p thì độ lệch chuẩn là lớn nhất?

Các Ứng dụng

4.15 Một cuộc điều tra của Viện Gallup được thực hiện vào tháng Giêng năm 1993 tiếp theo bài điều trần trước Quốc hội đã điều tra những người đã xem tất cả hay một phần bài điều trần này. ("Vấn đề nào nên là ưu tiên hàng đầu trong Chương trình Nghị sự 1994?, 1994"). Những người được phỏng vấn được hỏi vấn đề nào trong số bốn vấn đề sau đây nên nhận được sự ưu tiên cao nhất trong năm 1994: tội phạm, chăm sóc sức khỏe, thâm hụt ngân sách hay cải cách phúc lợi. Giải thích tại sao việc chọn mẫu này là và không phải là một thí nghiệm nhị thức.

4.16 Có một nỗi sợ hãi rằng việc quản lý các tổ chức chăm sóc sức khỏe (HMO) sẽ làm hạn chế các chọn lựa. Tuy nhiên, một nghiên cứu mới đây đã cho thấy rằng dân chúng mà phụ thuộc vào các HMO thì thỏa mãn với vấn đề chăm sóc sức khỏe hơn so với những thành viên thuộc các kế hoạch bảo hiểm y tế truyền thống (Braus, 1994). Nghiên cứu này đã so sánh 5.000 hộ gia đình đã đăng ký làm thành viên hoặc với một HMO, một kế hoạch bồi thường, hay một tổ chức cung cấp được ưa thích. Cuộc điều tra này ghi nhận rằng 85% số thành viên của HMO thỏa mãn với các chính sách bảo hiểm của họ. Giải thích tại sao cuộc điều tra về các thành viên của HMO là hay không phải là một thí nghiệm nhị thức.

4.17 Trong bài viết của mình, "High Court OKs Congress' Right to Regulate Cable Television (*Tòa án Tối cao Đồng ý về Quyền của Quốc hội trong việc Điều tiết Truyền hình Cáp*)", David Savage (1994) ghi nhận rằng 60% các hộ gia đình Hoa Kỳ có truyền hình cáp. Giả sử rằng $n = 4$ hộ gia đình được điều tra về việc liệu họ có hay không có truyền hình cáp. Giả định rằng con số 60% là chính xác trong việc trả lời các câu hỏi sau:

a. Xác suất để cho *x* bằng đúng với 4 là bao nhiêu?

b. Xác suất để cho *x* là 1 hay lớn hơn là bao nhiêu?

c. Xác suất để cho *x* bằng đúng với 1 là bao nhiêu?

4.18 Trong một nghiên cứu được tiến hành bởi *Business Marketing, Advertising Age,* và *USA Chicago, Inc.* ("Survey Said... (*Cuộc điều tra cho biết rằng...*)", 1994), các giám đốc điều hành, chuyên viên tiếp thị cao cấp, và giám đốc thông tin trên khắp đất nước đã được hỏi bằng cách nào mà siêu xa lộ thông tin sẽ tác động đến hoạt động kinh doanh, các thực tiễn tiếp thị và trách nhiệm cá nhân của họ. Khi các giám đốc điều hành được hỏi liệu họ có quan tâm đến siêu xa lộ thông tin quốc gia hay không, thì khoảng 50% trong số này trả lời là có. Giả định rằng con số này mang tính đại diện cho toàn bộ các giám đốc điều hành trên cả nước và rằng có 25 giám đốc điều hành được ngẫu nhiên chọn và hỏi rằng họ có quan tâm đến siêu xa lộ thông tin hay không.

a. Xác suất mà cả 25 người sẽ cho biết rằng họ có quan tâm đến siêu xa lộ thông tin là bao nhiêu?

b. Xác suất mà ít nhất 10 trong số 25 giám đốc điều hành sẽ cho biết rằng họ có quan tâm đến siêu xa lộ thông tin là bao nhiêu?

c. Xác suất mà đúng 10 giám đốc điều hành sẽ cho biết rằng họ có quan tâm đến siêu xa lộ thông tin là bao nhiêu?

4.19 Nhiều chủ doanh nghiệp nhận ra rằng một số người mà họ thuê mướn không phải là người hoặc điều mà họ tự cho mình là như thế. Việc rà soát những người xin việc làm mà giả mạo thông tin đã làm phát sinh một số lĩnh vực kinh doanh mới: dịch vụ kiểm tra phẩm chất của người xin việc. Giả định rằng bạn thuê mướn năm nhân viên mới trong tuần vừa qua và rằng xác xuất mà một nhân viên duy nhất làm giả mạo thông tin trong hồ sơ xin việc của anh/chị ta là 0,35. Xác suất mà ít nhất có một trong năm hồ sơ xin việc đã bị giả mạo thông tin là bao nhiêu? hai hồ sơ hay nhiều hơn sẽ là bao nhiêu?

4.20 Trong một cái nhìn toàn diện về thái độ của người Nhật và người Mỹ về nhau, người Nhật đã cho thấy niềm tự hào to lớn trong chất lượng các sản phẩm của họ. Tuy thế, họ cảm giác rằng Hoa Kỳ sẽ đóng một vai trò lớn hơn về cả sự lãnh đạo lẫn quyền lực kinh tế so với Nhật Bản trong những năm sắp đến. Cụ thể là, 71% số người Nhật cảm giác rằng sản phẩm của họ tốt hơn sản phẩm của người Mỹ, và 42% cảm nhận rằng Hoa Kỳ sẽ là cường quốc kinh tế số một thế giới trong thế kỷ tới ("How the Japanese See Themselves... and US (*Cách mà người Nhật tự đánh giá mình... và Hoa Kỳ*)", 1990). Giả định rằng 50 công dân Nhật được chọn lựa một cách ngẫu nhiên.

a. Phân phối xác suất của x, số người Nhật cho rằng sản phẩm của họ tốt hơn sản phẩm của người Mỹ là bao nhiêu?

b. Phân phối xác suất của x, số người Nhật cho rằng Hoa Kỳ sẽ là cường quốc kinh tế số một thế giới trong thế kỷ tới là bao nhiêu?

c. Tìm trung bình và độ lệch chuẩn của biến số ngẫu nhiên được mô tả trong phần (b).

d. Nếu chỉ có 5 trong số 50 công dân Nhật cho rằng Hoa Kỳ sẽ là cường quốc kinh tế số một thế giới trong thế kỷ tới, liệu có khả năng là con số 42% chính xác hay không? Hãy giải thích.

4.21 Nhiều công ty công ích đã bắt đầu thúc đẩy việc bảo toàn năng lượng bằng cách đề nghị các tỷ lệ chiết khấu cho những khách hàng mà giữ được mức sử dụng năng lượng thấp hơn các mức tiêu chuẩn trợ cấp được thiết lập nào đó. Ta biết rằng 70% cư dân của một thị trấn ở miền trung tây Hoa Kỳ đã giảm việc sử dụng điện năng của mình một cách hữu hiệu để đủ tiêu chuẩn được hưởng các tỷ lệ chiết khấu. Giả sử có năm cư dân từ thị trấn này được chọn ngẫu nhiên.

a. Xác suất mà cả năm đều đủ tiêu chuẩn được hưởng các tỷ lệ ưu đãi là bao nhiêu?

b. Xác suất mà có ít nhất bốn cư dân được hưởng các tỷ lệ ưu đãi là bao nhiêu?

4.22 Một máy móc được thiết kế để bơm đầy các lon một lượng soda là 12 oz. Sự thay đổi của các lần bơm làm cho mỗi lon có thể được bơm nhiều hơn hay ít hơn 12 oz soda. Nếu máy phân phối được cài đặt để cho 12 oz là lần bơm trung vị, thì xác suất của việc bơm dư là 0,5 và xác suất của việc bơm thiếu cũng là 0,5. Giả định rằng *n* lon được chọn từ dây chuyền sản xuất và số lon bơm thiếu được ghi nhận.

a. Xác suất mà tất cả các lon bị bơm thiếu khi $n = 3$ là bao nhiêu? khi $n = 5$ là bao nhiêu? khi $n = 10$ là bao nhiêu?

b. Nếu bạn là một người giám sát có trách nhiệm kiểm soát việc bơm soda vào các lon từ cái máy này, thì kết luận của bạn sẽ như thế nào nếu trên thực tế bạn quan sát thấy 3, rồi 5, và rồi 10 lon soda được chọn mẫu bị bơm thiếu?

c. Giải thích cách thức mà một lần chuyền chạy (một loạt các vật phẩm tương tự nhau) của các lon bị bơm thiếu có thể được sử dụng như là một biến số kiểm soát qui trình.

4.23 Theo một bào báo trên tờ *Los Angeles Times* (Horovitz, 1994) thì xúc xích nằm trong số các vật phẩm ít phổ biến nhất được bán tại các cửa tiệm bán thức ăn nhanh ở vùng Nam California. Một cuộc điều tra gồm $n = 600$ cư dân vùng Nam California được hỏi liệt kê vật phẩm thức ăn nhanh cuối cùng mà họ mua để dùng bữa. Trong số những cư dân được điều tra, 34% đã mua hamburger, 19% mua pizza và 11% mua thức ăn Mêhicô. Nếu như chỉ có 3,3% những người được điều tra mua xúc xích, thì chúng ta sẽ kỳ vọng con số người tiêu dùng thức ăn nhanh sẽ mua xúc xích rơi vào trong những giới hạn nào với xác suất xấp xỉ bằng 0,95 nếu như con số 3,3% này trên thực tế là chính xác?

4.3 PHÂN PHỐI XÁC SUẤT POISSON

Phân phối xác suất Poisson là một mô hình tốt cho sự phân phối tần suất tương đối của số lượng các sự kiện hiếm xảy ra mà xuất hiện trong một đơn vị thời gian, khoảng cách, không gian, và vân vân. Ví lý do này mà phân phối này thường được quản trị kinh doanh sử dụng để mô hình hóa sự phân phối tần suất tương đối của số lượng các vụ tai nạn công nghiệp theo một đơn vị thời gian (ví dụ như sự cố hạt nhân tại Đảo Ba Dặm), hay bởi các nhà quản lý nhân sự để mô hình hóa sự phân phối tần suất tương đối cho số lượng tại nạn của nhân viên hay số lượng những lần yêu sách đòi bảo hiểm theo một đơn vị thời gian. Phân phối xác suất Poisson cũng có thể, trong một số trường hợp, cung cấp một mô hình tốt cho sự phân phối tần suất tương đối của số lượng đối tượng đến theo một đơn vị thời gian tại một điểm phục vụ (ví dụ, số lượng đơn hàng nhận được tại một nhà máy sản xuất hay số lượng khách hàng tại một cơ sở tiện ích dịch vụ, một quầy tính tiền ở siêu thị, v.v). Công thức cho phân phối Poisson được thể hiện như sau:

Phân phối Xác suất Poisson

$$p(x) = \frac{\mu^x e^{-\mu}}{x!}, \quad x = 0,1,2,\ldots,\infty$$

trong đó x = số lượng các sự kiện hiếm xảy ra theo một thời vị thời gian, khoảng cách, không gian và vân vân và e = 2.7182818...

Trung bình: ký hiệu μ mà xuất hiện trong $p(x)$

Phương sai: $\sigma^2 = \mu$

Độ lệch chuẩn: $\sigma = \sqrt{\mu}$

Giá trị của $e^{-\mu}$ có thể được tìm thấy bằng cách sử dụng hầu hết các máy tính bỏ túi.

Lưu ý rằng, trên thực tế thì *x* thường nhỏ; về mặt lý thuyết thì nó có thể lớn vượt quá tất cả các giới hạn. Như vậy biến số ngẫu nhiên Poisson là một ví dụ về một biến số ngẫu nhiên rời rạc mà có thể có một số lượng các giá trị lớn vô cùng (nhưng có thể đếm được). Phân phối Poisson là duy nhất trong số các phân phối rời rạc ở chỗ $\mu = \sigma^2$ – nghĩa là, trung bình bằng với phương sai.

Phân phối xác suất Poisson cũng có thể được sử dụng để ước lượng xấp xỉ phân phối nhị thức khi *n* lớn và *p* nhỏ[+] và khi trung bình nhị thức là nhỏ hơn 7. Trong trường hợp này, giá trị của 1- *p* sẽ tiến gần đến 1, trung bình nhị thức *np* sẽ xấp xỉ bằng với phương sai nhị thức *np* (1-*p*), và các xác suất Poisson với $\mu = np$ sẽ xấp xỉ gần với các phân phối nhị thức cho *n* và *p* đã biết.

Chúng ta sẽ minh họa hai loại hình ứng dụng này trong các ví dụ tiếp theo đây. Những ví dụ khác sẽ được đề xuất qua các bài tập.

VÍ DỤ **4.7** Những vụ thương tật nghiêm trọng của công nhân tại một công ty sản xuất thép bình quân là 2,7 vụ một năm. Đã biết các điều kiện an toàn tại nhà máy vẫn giữ nguyên trong năm tới, xác suất mà con số các vụ thương tật nghiêm trọng sẽ nhỏ hơn 2 là bao nhiêu?

Lời giải Sự kiện mà có ít hơn hai vụ thương tật nghiêm trọng sẽ xảy ra là sự kiện rằng *x* = 0 hay 1. Do vậy,

$$P(x<2) = p(0) + p(1) \text{ trong đó } p(x) = \frac{(2{,}0)^x e^{-2{,}7}}{x!}$$

Thay thế vào phương trình cho $p(x)$ với $e^{-2{,}7} = 0{,}067206$, chúng ta có được

$$\begin{aligned} P(x<2) = p(0) + p(1) &= \frac{(2{,}7)^0(0{,}067206)}{0!} + \frac{(2{,}7)^1(0{,}067206)}{1!} \\ &= (0{,}067206) + (2{,}7)90{,}067206) \\ &= 0{,}249 \end{aligned}$$

(Nhắc lại rằng 0!=1). Vì vậy, xác suất để cho có ít hơn hai vụ thương tật nghiêm trọng đối với công nhân sẽ xảy ra trong năm tới tại nhà máy sản xuất thép là 0,249.

Để cho thuận tiện, chúng tôi cung cấp trong Bảng 2 của Phụ lục II các số tổng một phần, $\sum_{x=0}^{a} p(x)$, cho phân phối xác suất Poisson cho các giá trị của μ từ 0,25 đến 5,0 trong các bước của 0,25. Bảng này được xây dựng theo cùng cách thức như bảng các số tổng một phần đối với phân phối xác suất nhị thức, Bảng 1 của Phụ lục II. Ví dụ sau đây minh họa việc sử dụng Bảng 2 và cũng biểu diễn việc sử dụng phân phối xác suất Poisson để ước lượng xấp xỉ phân phối xác suất nhị thức.

VÍ DỤ **4.8** Giả sử rằng bạn có một thí nghiệm nhị thức với *n* = 25 và *p* = 0,1. Hãy tìm ra giá trị chính xác của $P(x \le 3)$ bằng cách sử dụng bảng các số tổng một phần cho phân phối xác suất nhị thức, Bảng 1 của Phụ lục II. Sau đó tìm số tổng một phần tương ứng bằng cách sử dụng ước lượng xấp xỉ Poission trong Bảng 2 của Phụ lục II. So sánh các giá trị chính xác và xấp xỉ cho $P(x \le 3)$.

[+] Nếu *p* gần bằng 1, hãy thay đổi định nghĩa của bản về thành công và thất bại để cho *p* gần bằng zêrô.

Lời giải Từ Bảng 1 trong Phụ lục II, giá trị chính xác của $P(x \le 3)$ là $\sum_{x=0}^{3} p(x) = 0{,}764.$ Số tổng Poisson tương ứng, cho $\mu = np = (25)(0{,}1) = 2{,}5,$ được cho trong Bảng 2 của Phụ lục II, là $P(x \le 3) = \sum_{x=0}^{3} p(x) = 0{,}758.$ So sánh các giá trị chính xác và xấp xỉ của $P(x \le 3)$, chúng ta thấy rằng giá trị xấp xỉ là rất tốt. Nó chỉ khác với giá trị chính xác có 0,006.

Minitab và Excel sẽ cho phép bạn tạo ra hoặc các xác suất Poisson riêng lẻ hay cộng dồn. Trong Minitab, sử dụng **Calc → Probability Distributions → Poisson,** bằng cách sử dụng hoặc "Probability" hoặc "Cumulative Probability", và chỉ định các giá trị cần thiết cho *x*. Nhập vào một giá trị cho μ vào hộp có ghi chú "Mean", và bấm **OK**. Trong Excel, sử dụng **Insert → Function → POISSON,** bằng cách chọn dãy những ô có chứa các giá trị *x*, giá trị cho μ, và lựa chọn của bạn về xác suất cộng dồn hay riêng lẻ. Để biết về sự giải thích chi tiết hơn, tham khảo **Sử dụng Excel** trong Phụ lục V. Các xác suất nhị thức cộng dồn cho $n = 25$ và $p = 0{,}1$, cùng với các phân phối Poisson cộng dồn cho $\mu = 2{,}5$ được cho trong Bảng 4.3. Lưu ý rằng các ước lượng xấp xỉ Poisson cho những xác suất nhị thức thực sự là hoàn toàn chính xác trong trường hợp này.

BẢNG 4.3 **Bảng in từ Minitab về các xác suất nhị thức và Poisson**

NHỊ THỨC VỚI N = 25 P = 0,100000		POISSON VỚI TRUNG BÌNH = 2.500	
K	P (X ≤ K)	K	P (X ≤ K)
0	0,0718	0	0,0821
1	0,2712	1	0,2873
2	0,5371	2	0,5438
3	0,7636	3	0,7576
4	0,9020	4	0,8912
5	0,9666	5	0,9580
6	0,9905	6	0,9858
7	0,9977	7	0,9958
8	0,9995	8	0,9989
9	0,9999	9	0,9997
10	1,0000	10	0,9999
		11	1,0000

Bài tập

Các Kỹ thuật Cơ bản

4.24 Giả định *x* là một biến số ngẫu nhiên Poisson với $\mu = 1{,}2$. Hãy tìm kết quả cho:

a $p(0)$

b $p(1)$

c $P(x \le 2)$

d $P(x > 1)$

4.25 Giả định rằng *x* là một biến số ngẫu nhiên Poisson với $\mu = 2$. Hãy tìm kết quả cho:

a $p(0)$

b $P(x > 1)$

c $P(x < 2)$

4.26 Sử dụng Bảng 2 trong Phụ lục II để tìm $p(x)$ cho một phân phối xác suất Poisson với $\mu = 1$ và x = 0, 1, 2, 3, 4,...Sau đó vẽ đồ thị $p(x)$.

4.27 Lặp lại các chỉ dẫn của Bài tập 4.26 cho $\mu = 3$. Lưu ý cách thức mà phân phối có xu hướng trở nên có hình dạng gò nhiều hơn khi μ tăng lên.

4.28 Sử dụng Bảng 2 trong Phụ lục II để tìm kết quả cho:

a $P(x \leq 2)$ khi $\mu = 3$

b $P(x \geq 1)$ khi $\mu = 1$

c $P(x = 2)$ khi $\mu = 2$ [*Gợi ý:* $p(x) = \sum_{x=0}^{2} p(x) - \sum_{x=0}^{1} p(x)$.]

4.29 Một thí nghiệm nhị thức có n = 20 và p = 0,2.

a Sử dụng Bảng 1 để tìm ra giá trị chính xác của $p(2)$.

b Sử dụng Bảng 2 để tìm ra ước lượng xấp xỉ Poisson cho $p(2)$.

Các Ứng dụng

4.30 Giả định rằng một hệ thống tuần tra cảnh sát ngẫu nhiên được thiết kế để cho một nhân viên cảnh sát tuần tra có thể viếng thăm một nơi đã biết theo nhịp đi x = 0, 1, 2, 3,... lần mỗi quãng thời gian nửa tiếng và rằng hệ thống này được xếp đặt sao cho anh ta viếng thăm mỗi nơi trung bình một lần trong mỗi quãng thời gian. Giả định rằng x sở hữu xấp xỉ một phân phối xác suất Poisson. Hãy tính xác suất mà nhân viên cảnh sát tuần tra sẽ bỏ lỡ một nơi đã biết trong suốt quãng thời gian nửa tiếng. Xác suất mà anh ta sẽ viếng thăm nó một lần là bao nhiêu? hai lần? ít nhất một lần?

4.31 Tại nạn tại một nhà máy công nghiệp cụ thể bình quân là 3,5 vụ mỗi tuần.

a Xác suất mà không có tai nạn nào xảy ra trong một tuần đã biết là bao nhiêu?

b Có khả năng rằng số lượng tai nạn mỗi tuần sẽ vượt quá 7 không? Hãy giải thích.

c Nếu số lượng tai nạn trong một tuần cụ thể là bằng với 9, liệu bạn còn tin rằng $\mu = 3{,}5$ không? Hãy giải thích.

4.32 Số lượng x, mỗi tuần, về doanh số thiết bị lớn đo chấn động trái đất bán cho một công ty thiết bị xây dựng sở hữu một phân phối xác suất Poisson với trung bình bằng 4.

a Xác suất mà con số thiết bị đo chấn động trái đất bán được mỗi tuần là bằng với 1 là bao nhiêu? ít hơn hay bằng 1?

b Liệu có khả năng rằng x sẽ vượt quá 9 không? Giải thích.

4.33 Chủ sở hữu duy nhất của một văn phòng bất động sản dân cư lưu ý rằng, tính trung bình thì yêu cầu tìm hiểu thông tin qua điện thoại đến một cách ngẫu nhiên và độc lập ở mức bốn lần mỗi ngày làm việc 8 tiếng. Bởi vì chủ sở hữu bất động sản thường đi ra ngoài với các thân chủ của mình, bà ta không thể trả lời ngay lập tức cho một số yêu cầu tìm hiểu qua điện thoại tại văn phòng của mình.

a. Xác suất mà không có yêu cầu tìm hiểu thông tin qua điện thoại nào đến trong 2 giờ vắng mặt tại văn phòng trong suốt một ngày làm việc điển hình 8 tiếng là bao nhiêu?

b. Xác suất mà sẽ có ít nhất 5 yêu cầu tìm hiểu thông tin qua điện thoại trong suốt một ngày làm việc điển hình 8 tiếng là bao nhiêu?

4.34 Số lượng *x* người đi vào một đơn vị chăm sóc sức khỏe chuyên sâu tại một bệnh viện cụ thể vào bất cứ một ngày nào sở hữu một phân phối xác suất Poisson với trung bình bằng với 5 người mỗi ngày.

a. Xác suất mà số người đi vào một đơn vị chăm sóc sức khỏe chuyên sâu vào một ngày cụ thể là bằng với 2 là bao nhiêu? ít hơn hay bằng 2 là bao nhiêu?

b. Liệu có khả năng rằng x sẽ vượt quá 10? Giải thích.

4.4 CÁC PHÂN PHỐI XÁC SUẤT RỜI RẠC KHÁC (TÙY CHỌN)

Trong Ví dụ 3.1 và Bài tập 3.51, thí nghiệm bao gồm việc tung đồng tiền và quan sát *x*, số lần xuất hiện mặt ngửa của đồng tiền. Đây là một ví dụ của một thí nghiệm tổng quát hơn mà trong đó biến số ngẫu nhiên có thể nhận các giá trị $x = 1, 2, ..., N$ với xác suất như nhau khi đã biết rằng $p(x) = 1/N$. Phân phối tạo ra cho *x* được gọi là **phân phối xác suất đồng nhất rời rạc** bởi vì biểu đồ xác suất tạo ra có chiều cao đồng nhất.

Một phân phối xác suất rời rạc thứ hai, mà tương tự với phân phối nhị thức, xảy ra nếu bạn chọn một mẫu ngẫu nhiên của *n* khách hàng từ một tổng thể có *N* khách hàng. Con số *x* khách hàng ưa thích một sản phẩm cụ thể hơn sẽ sở hữu một phân phối xác suất nhị thức khi cỡ mẫu *n* là tương đối nhỏ so với con số *N* khách hàng trong tổng thể (xem Ví dụ 4.1). Khi *n* là tương đối lớn so với *N* (như trong Ví dụ 4.2), thì con số *x* ưa thích sản phẩm đó hơn sở hữu một **phân phối xác suất siêu bội.** Công thức của nó được trình bày trong phần sau đây.

Phân phối Xác suất Siêu bội

$$p(x) = \frac{C_x^r C_{n-x}^{N-r}}{C_n^N}$$

trong đó

N = số lượng các phần tử trong tổng thể

r = số lượng các phần tử sở hữu một số đặc trưng đặc biệt, ví dụ, số lượng người ưa thích một sản phẩm cụ thể hơn

n = số lượng các phần tư trong mẫu chọn

Trung bình: $\mu = n\left(\frac{r}{N}\right)$

Phương sai: $\sigma^2 = n\left(\frac{r}{N}\right)\left(\frac{N-r}{N}\right)\left(\frac{N-n}{N-1}\right)$

Độ lệch chuẩn: $\sigma = \sqrt{\sigma^2}$

Khi cỡ mẫu *n* là tương đối nhỏ so với cỡ tổng thể *N* và *n/N* là ít hơn 0,05, thì các xác suất siêu bội có thể được ước lượng xấp xỉ bởi phân phối nhị thức với $p = r/N$.

VÍ DỤ **4.9** Một hội thẩm đoàn bao gồm 20 người, 2 trong số này là người Mỹ bản địa. Nếu 3 người được chọn ngẫu nhiên từ hội thẩm đoàn này, xác suất mà hai người là người Mỹ bản địa là bao nhiêu?

Lời giải Đối với ví dụ này,

$N = 20$, $n = 3$

$r = 2$ (người Mỹ bản địa)

x = số người Mỹ bản địa trong sự chọn lựa

Sau đó,

$$p(x) = \frac{C_x^r C_{n-x}^{N-r}}{C_n^N}$$

và

$$p(2) = \frac{C_2^2 C_{3-2}^{20-2}}{C_3^{20}}$$

trong đó

$$C_2^2 = \frac{2!}{2!0!} = 1, \quad C_{3-2}^{20-2} = C_1^{18} = \frac{18!}{1!17!} = 18$$

và

$$C_3^{20} = \frac{20!}{3!17!} = \frac{(20)(19)(18)}{6} = 1140$$

Xác suất của việc có hai người Mỹ bản địa trong mẫu có cỡ mẫu $n = 3$ là

$$p(2) = \frac{(1)(18)}{1140} = 0{,}016$$

Bài tập

Các Kỹ thuật Cơ bản

4.35 **a** Tính toán $p(x)$, trong đó x có một phân phối xác suất siêu bội với N =10, n = 2, r = 3, và x = 0, 1, 2.

b Vẽ đồ thị $p(x)$.

4.36 **a** Tính toán $p(x)$, trong đó x có một phân phối xác suất siêu bội với N =10, n = 3, r = 3, và x = 0, 1, 2, 3.

b Vẽ đồ thị $p(x)$.

4.37 Tìm trung bình và độ lệch chuẩn cho biến số ngẫu nhiên x được mô tả trong Bài tập 4.36. Xác suất mà x nằm trong khoảng $(\mu \pm 2\sigma)$ là bao nhiêu?

Các Ứng dụng

4.38 Một vấn đề mà các giám đốc nhân sự gặp phải cũng như những giám đốc khác phải đối mặt là sự chọn lựa cái tốt nhất trong một tập hợp hữu hạn của các yếu tố được thể hiện trong tình huống sau: Từ một nhóm 20 Tiến sĩ kỹ sư, 10 người được chọn làm việc. Xác suất mà 10 người được chọn bao gồm 5 kỹ sư tốt nhất trong nhóm 20 người là bao nhiêu?

4.39 Một sản phẩm công nghiệp cụ thể được chuyên chở và giao hàng theo từng lô 20 sản phẩm. Việc kiểm tra để quyết định liệu một mẫu hàng có lỗi hay không là hết sức tốn kém; vì thế mà nhà sản xuất chọn mẫu sản phẩm hơn là sử dụng phương án kiểm tra toàn bộ sản phẩm sản xuất ra. Một phương án chọn mẫu được thiết kế nhằm giảm thiểu số sản phẩm bị lỗi được giao cho khách hàng đòi hỏi việc chọn 5 mẫu hàng từ mỗi lô hàng và từ chối lô hàng nào có nhiều hơn một sản phẩm bị lỗi quan sát thấy được. (Nếu bị từ chối, thì từng sản phẩm trong lô hàng đó sẽ bị kiểm tra). Nếu một lô hàng có bốn sản phẩm bị lỗi, thì xác suất mà lô hàng đó bị từ chối là bao nhiêu?

4.40 Texaco trở thành công ty dầu mỏ lớn gần đây nhất phải cắt giảm lực lượng lao động và chấm dứt khai thác một số mỏ dầu ở Mỹ của mình (Craig, 1994). Exxon và Mobil gần đây đã tuyên bố việc tái cấu trúc tương tự nhằm gia tăng lợi nhuận. Giả định rằng ba trong số mười công ty lọc dầu hàng đầu của Mỹ trên thực tế đang tiến hành việc tái cấu trúc công ty. Nếu một phóng viên của tờ *USA Today* phỏng vấn giám đốc điều hành của bốn công ty lọc dầu được chọn lựa ngẫu nhiên, hãy tính toán các xác suất sau.

a. Chọn lựa của cô ta bao gồm tất cả 3 giám đốc điều hành mà công ty của họ được tái cấu trúc gần đây.

b. Chọn lựa của cô ta bao gồm không có giám đốc điều hành nào mà công ty của họ được tái cấu trúc gần đây.

c. Chọn lựa của cô ta bao gồm ít nhất một trong số các giám đốc điều hành mà công ty của họ được tái cấu trúc gần đây.

QUAY TRỞ LẠI NGHIÊN CỨU ĐIỂN HÌNH

4.5 TIẾP TỤC VỀ QUYỀN LÁI XE

Như chúng tôi đã gợi ý trong nghiên cứu điển hình, sự ước tính tỷ phần người Mỹ trưởng thành mà ưa thích hơn về một bài kiểm tra mang tính bắt buộc mỗi ba năm cho các tài xế trên 65 tuổi tùy thuộc vào phân phối xác suất của *x*, số người trong cuộc điều tra này mà ủng hộ các bài kiểm tra mang tính bắt buộc cho các công dân lớn tuổi. Bởi vì số người được liên hệ trong cuộc điều tra cấu thành một mẫu ngẫu nhiên từ một con số lớn dân chúng, cho nên trên thực tế *x* sẽ sở hữu một phân phối xác suất nhị thức.

Giả sử rằng con số 70% thật sự là giá trị chính xác của *p*. Khi đã biết một mẫu có qui mô 1.003, thì số người trong mẫu này mà cho thấy rằng họ ủng hộ các bài kiểm tra mang tính bắt buộc đối với các tài xế trên 65 tuổi sẽ sở hữu một phân phối nhị thức với trung bình và độ lệch chuẩn bằng với

$$\mu = np = (1003)(0,7) = 702,1$$

và

$$\sigma = \sqrt{npq} = \sqrt{(1003)(0,7)(0,3)} = 14,51$$

Như vậy, nếu *p* trên thực tế là 0,7 thì theo Qui tắc Thực chứng với xác suất đúng 95%, chúng ta sẽ kỳ vọng rằng con số trong mẫu này nằm trong khoảng

$$\mu \pm 2\sigma = 702,1 \pm 2(14,51) = (673,08;\ 731,12)$$

hay từ 674 đến 731 người trả lời.

Nếu trên thực tế *p* là 0,7 và chúng ta quan sát thấy $x = 600$ trong số 1.003 người trả lời mà đồng ý với các bài kiểm tra mang tính bắt buộc dành cho tài xế trên 65 tuổi, thì chúng ta sẽ nghiêng về việc nghĩ rằng giá trị của $p = 0,7$ là quá lớn và rằng tỷ lệ thực tế số người trả lời ủng hộ gần hơn với $p = 0,6$. Khi chúng ta nhận ra rằng giá trị thực tế của *p* không phải là 0,7 mà thay vào đó là rằng *ước lượng mẫu của p* là 0,7, chúng ta có lẽ muốn tìm ra những giới hạn mà giá trị thật sự của *p* nằm trong đó. Bởi vì con số những câu trả lời ủng hộ nằm trong khoảng 674 đến 731 ở mức xấp xỉ 95% toàn thời gian, chúng ta sẽ ước lượng rằng *p* nằm trong khoảng

$$\left(\frac{674}{1003}, \frac{731}{1003}\right) = (0,672;\ 0,729)$$

hay từ 0,672 đến 0,729.

4.6 TÓM TẮT

Sự quen thuộc với các phân phối xác suất rời rạc và những đặc tính của các thí nghiệm mà tạo ra những phân phối này là hết sức hữu ích. Thay vì cứ lặp đi lặp lại việc giải quyết cùng một vấn đề xác suất từ các nguyên lý đầu tiên (như đã được thực hiện trong Chương 3), thì bạn chỉ cần nhận thức được loại hình biến số ngẫu nhiên và sau đó thay thế vào trong công thức tính phân phối xác suất của chúng.

Nhiều phân phối xác suất rời rạc hữu ích được trình bày trong chương này: phân phối nhị thức, phân phối Poisson và phân phối siêu bội. Những phân phối xác suất này cho phép chúng ta có thể tính toán các xác suất đi cùng với những sự kiện thuộc lĩnh vực quan tâm của các ngành khoa học, kinh doanh và tiếp thị.

Phân phối xác suất nhị thức cho phép chúng ta tính toán xác suất của *x* số thành công trong một loạt *n* lần thử độc lập giống nhau, mà ở đó xác suất của một thành công trong một lần thử duy nhất là bằng với *p*. Thí nghiệm nhị thức là một mô hình tuyệt vời cho nhiều tình huống chọn mẫu, đặc biệt là các cuộc điều tra mà tạo ra các loại hình dữ liệu "có" và "không".

Phân phối xác suất Poisson là quan trọng bởi vì nó có thể được sử dụng để ước lượng xấp xỉ một số xác suất nhị thức nào đó khi *n* lớn và *p* nhỏ. Vì thế, nó có thể làm giảm đi rất nhiều các phép tính có liên quan đến việc tính toán các xác suất nhị thức. Hơn nữa, phân phối xác suất Poisson tự thân nó là quan trọng. Nó cung cấp cho ta một mô hình xác suất tuyệt vời cho nhiều sự hiện diện các sự kiện hiếm khi xảy ra trong thời gian và không gian.

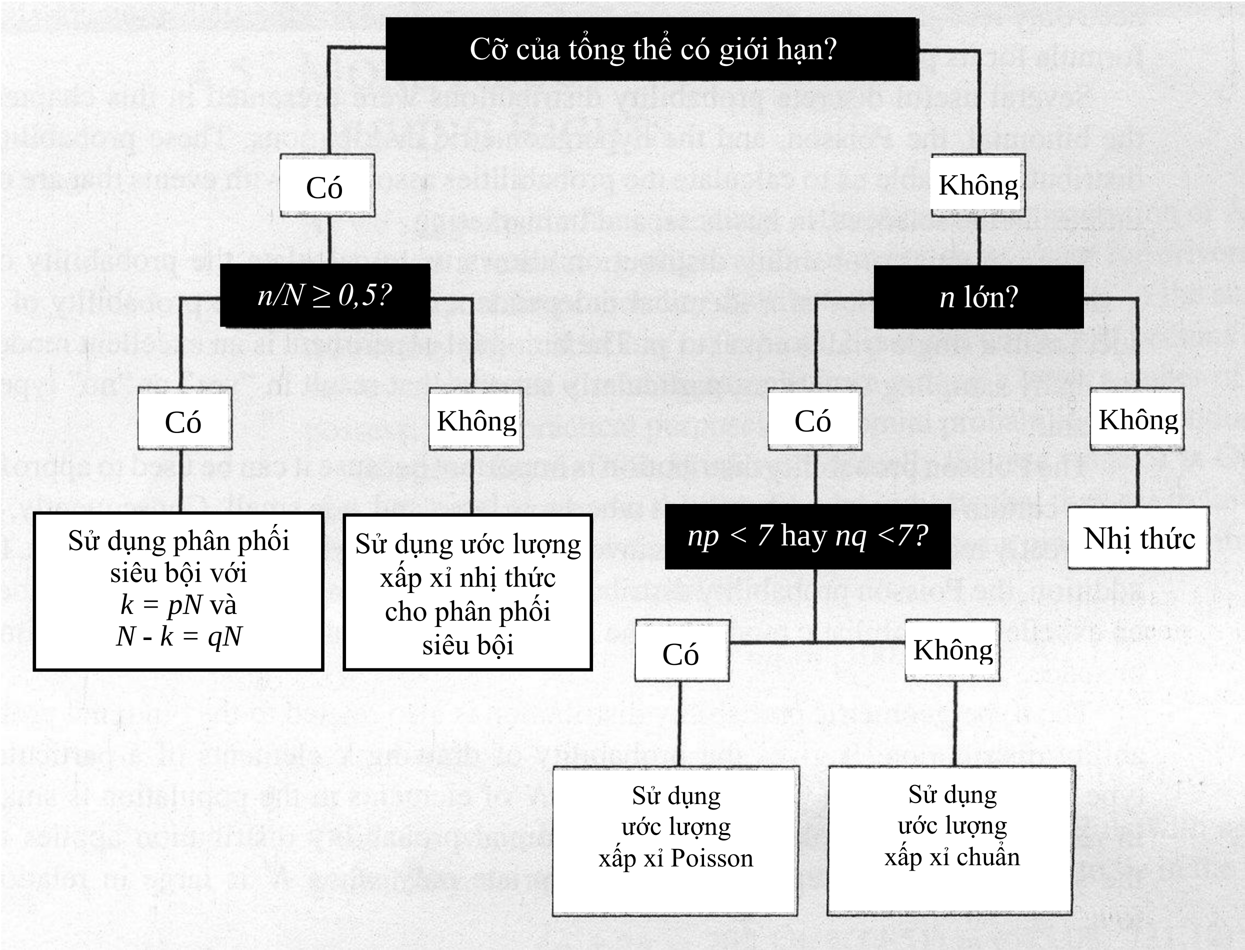
Phân phối xác suất siêu bội cũng có liên quan đến phân phối xác suất nhị thức. Nó tạo cho xác suất của việc rút ra *x* yếu tố của một loại hình đặc thù từ một tổng thể mà ở đó số lượng *N* yếu tố trong tổng thể này là nhỏ so với cỡ mẫu *n*. Phân phối xác suất nhị thức áp dụng được cho cùng tình huống này ngoại trừ nó chỉ phù hợp khi *N* là tương đối lớn so với *n*.

Những ghi nhớ cho việc giải quyết vấn đề

1 Nếu một biến số ngẫu nhiên là số lượng những lần xảy ra của một sự kiện đặc biệt trong một đơn vị *thời gian* biết trước mà qua đó số lượng trung bình của những lần xảy ra theo mỗi đơn vị thời gian hay không gian là μ, thì biến số ngẫu nhiên đó có một phân phối Poisson.

2 Giả định rằng một mẫu ngẫu nhiên gồm *n* vật phẩm được chọn mà không có sự thay thế từ tổng thể bao gồm *N* vật phẩm mà trong đó một tỷ lệ *p* của các vật phẩm này sở hữu một đặc trưng riêng biệt và $q = 1 - p$ thì không. Sơ đồ trong Hình 4.3 được cung cấp nhằm giúp bạn quyết định phân phối nào là phù hợp cho các tình huống trình bày.

HÌNH 4.3 **Cây quyết định**



Bài tập thêm

Các bài tập có đánh dấu (*) là tùy chọn

4.41 Hãy liệt kê năm đặc trưng xác định của thí nghiệm nhị thức.

4.42 Một đồng xu cân bằng được tung lên ba lần. Cho x bằng với số lần đếm quan sát được.

a. Sử dụng công thức cho phân phối xác suất nhị thức để tính toán các xác suất đi cùng với x = 0, 1, 2 và 3.

b. Xây dựng một phân phối xác suất tương tự như các phân phối được thể hiện trong Hình 4.1.

c. Tìm giá trị kỳ vọng và độ lệch chuẩn của x, bằng cách sử dụng công thức

d. $E(x) = np \qquad \sigma = \sqrt{npq}$

e. Sử dụng phân phối xác suất mà bạn suy ra từ (b), hãy tìm tỷ lệ đại lượng của tổng thể nằm bên trong một độ lệch chuẩn tính từ số trung bình. Lặp lại cho hai độ lệch chuẩn. Bằng cách nào mà các kết quả của bạn nhất quán với Định lý Tchebysheff và Qui tắc Thực chứng?

4.43 Sử dụng Bảng 1 trong Phụ lục II để tìm số tổng một phần $\sum_{x=0}^{a} p(x)$ cho một biến số ngẫu nhiên nhị thức x với các đặc trưng như sau.

a $n = 10, \quad p = 0{,}7, \quad a = 8$

b $n = 15, \quad p = 0{,}05, \quad a = 1$

c $n = 20, \quad p = 0{,}9, \quad a = 14$

4.44 Sử dụng Bảng 1 trong Phụ lục II để tìm $p(x)$ cho một biến số ngẫu nhiên nhị thức x với các đặc trưng sau đây

a $n = 10, \quad p = 0{,}6, \quad x = 6$

b $n = 15, \quad p = 0{,}5, \quad x = 5$

c $n = 20, \quad p = 0{,}2, \quad x = 3$

4.45 Sử dụng Bảng 1 trong Phụ lục II để tìm $\sum_{x=a}^{b} p(x)$ cho một biến số ngẫu nhiên nhị thức x với các đặc trưng sau đây

a $n = 10, \quad p = 0{,}1, \quad a = 1, \quad b = 10$

b $n = 10, \quad p = 0{,}8, \quad a = 7, \quad b = 9$

c $n = 15, \quad p = 0{,}4, \quad a = 4, \quad b = 15$

4.46 Sử dụng Bảng 1 trong Phụ lục II để tìm các xác suất sau cho biến số ngẫu nhiên nhị thức x.

a $P(x \le 1)$ cho $n = 5,\ p = 0{,}2$

b $P(x > 1)$ cho $n = 5,\ p = 0{,}2$

c $P(x < 1)$ cho $n = 5,\ p = 0{,}2$

d $P(x = 1)$ cho $n = 5,\ p = 0{,}2$

e $P(x = 1)$ cho $n = 10,\ p = 0{,}2$

4.47 Sử dụng Bảng 1 trong Phụ lục II để tìm các xác suất sau cho biến số ngẫu nhiên nhị thức x.

a $P(x \le 4)$ cho $n = 10,\ p = 0{,}5$

b $P(x < 4)$ cho $n = 10,\ p = 0{,}5$

c $P(x > 4)$ cho $n = 10,\ p = 0{,}5$

d $P(x = 4)$ cho $n = 10,\ p = 0{,}5$

4.48 Một nghiên cứu gần đây về các chuyến tàu chở những người sống ở ngoại ô vào làm việc tại thành phố cho thấy rằng các chuyến tàu này chạy trễ hơn 35 phút với xác suất bằng 0,5. Nếu chúng ta đang sử dụng một tàu chở những người ở ngoại ô vào làm việc tại thành phố mà thể hiện những đặc trưng này và chúng ta chọn ngẫu nhiên năm ngày trong năm vừa qua, thì xác suất mà chuyến tàu này luôn luôn bị trễ là bao nhiêu? Xác suất mà chuyến tàu này bị trễ nhiều hơn ba lần trong số năm lần là bao nhiêu?

4.49 Một nhà kho chứa mười máy in, bốn trong số này bị lỗi. Một công ty chọn năm trong số các máy in này một cách ngẫu nghiên, nghĩ rằng tất cả đều trong điều kiện hoạt động tốt. Xác suất mà tất cả năm máy in đều không có lỗi là bao nhiêu?

4.50 Một công đoàn cho rằng 45 trong số 80 nhân viên của một công ty ủng hộ việc thành lập công đoàn. Giả định rằng công đoàn đó đúng, và người quản lý nhà máy một cách không chính thức đã chọn mẫu ý kiến của 20 nhân viên.

a. Giá trị kỳ vọng của con số *x* nhân viên trong mẫu này mà sẽ ủng hộ việc lập công đoàn là bao nhiêu?

b. Tìm độ lệch chuẩn của *x*.

c. Nếu công đoàn đúng, có khả năng xảy ra là có ít hơn 9 nhân viên trong mẫu này sẽ ủng hộ việc lập công đoàn. Giải thích.

4.51 Một nghiên cứu của chính phủ Hoa Kỳ về các cuộc gọi điện thoại của các nhân viên cho thấy rằng cứ một trong ba cuộc gọi là cho các mục đích ngoài công việc (*New York Times,* ngày 23 tháng Sáu năm 1986). Giả sử rằng bạn là một nhân viên chính phủ và rằng ba trong số mỗi mười cuộc gọi điện thoại của bạn là cho các lý do cá nhân. Chính phủ đã chọn ngẫu nhiên mười số điện thoại mà bạn đã gọi đến.

a. Xác suất mà không nhiều hơn một cuộc gọi là vì các lý do cá nhân là bao nhiêu?

b. Xác suất mà có nhiều hơn năm một cuộc gọi là vì các lý do cá nhân là bao nhiêu?

c. Xác suất mà có đúng ba cuộc gọi là vì các lý do cá nhân là bao nhiêu?

4.52 Trong Bài tập 4.51, chúng ta đã lưu ý rằng một nghiên cứu của chính phủ cho thấy rằng cứ một trong ba cuộc gọi điện thoại của các nhân viên chính phủ là cho các mục đích ngoài công việc. Giả sử rằng một cơ quan chính phủ thực hiện 10 triệu cuộc gọi mỗi năm và rằng mỗi cuộc gọi tốn kém tiền của người trả thuế là 1,50 USD cho việc sử dụng thiết bị liên lạc và thời gian của nhân viên.

a. Con số kỳ vọng của các cuộc gọi cho mục đích ngoài công việc là bao nhiêu?

b. Độ lệch chuẩn là bao nhiêu?

c. Liệu con số *x* của các cuộc gọi ngoài công việc có thể ít hơn 3,3 triệu không? Giải thích.

4.53 Công ty Honda đã gia tăng hoạt động tại khu vực Bắc Mỹ của mình trong một vài năm qua, tạo thêm công ăn việc làm, gia tăng sản xuất, và giới thiệu các mẫu xe hơi mới tại các nhà máy ở Bắc Mỹ của mình. Hiện nay, khoảng 60% tất cả các loại xe hơi Honda bán tại Bắc Mỹ cũng được làm tại đây (Prodis, 1994). Giả sử rằng 15 xe Honda được mua tại Bắc Mỹ được chọn ngẫu nhiên và kiểm tra.

a. Xác suất mà 8 hay ít hơn các xe Honda được sản xuất tại Bắc Mỹ là bao nhiêu?

b. Xác suất mà từ 8 đến 12 xe Honda được sản xuất ra tại Bắc Mỹ là bao nhiêu?

c. Xác suất mà có đúng 10 xe Honda được sản xuất tại Bắc Mỹ là bao nhiêu?

4.54 Tham khảo lại Bài tập 4.53. Một công ty nhỏ điều tra các nhân viên của mình và nhận thấy rằng 15 người trong số họ lái xe hơi Honda. Trong số 15 xe Honda này, 8 xe được sản xuất ra tại Bắc Mỹ. Nếu năm nhân viên được chọn ngẫu nhiên để cho biết ý kiến về hiệu suất hoạt động của xe Honda, thì xác suất mà cả năm người đều có xe được sản xuất tại Bắc Mỹ là bao nhiêu?

4.55 Bạn biết rằng 90% những người mua một chiếc tivi màu sẽ không có yêu sách nào về bảo hành trong suốt thời hạn bảo hành. Giả sử rằng có 20 khách hàng mà mỗi người trong số họ mua một chiếc tivi từ một đại lý bán đồ điện gia dụng nhất định. Xác suất mà có ít nhất hai trong số 20 khách hàng này sẽ yêu cầu được bảo hành là bao nhiêu?

4.56 Giả sử rằng bạn biết một trong mười cuốn sách giáo khoa cấp đại học là một thành công tài chính nổi bật. Một nhà xuất bản đã chọn mười cuốn sách giáo khoa mới để xuất bản.

a. Xác suất mà có đúng một cuốn sách sẽ là một thành công tài chính đáng kể là bao nhiêu?

b. Xác suất mà có ít nhất một cuốn sách sẽ là một thành công tài chính đáng lưu ý là bao nhiêu?

c. Xác suất mà có ít nhất hai cuốn sách sẽ là những thành công tài chính nổi bật là bao nhiêu?

4.57 Tỷ lệ các hộ dân cư tại một thành phố nhỏ mà được sưởi ấm bằng khí thiên nhiên là xấp xỉ 0,2. Một khu phố được chọn ngẫu nhiên trong giới hạn thành phố này có 20 hộ gia đình đang sinh sống. Giả sử rằng các đặc trưng của một thí nghiệm nhị thức được thỏa mãn.

a. Tìm xác suất mà không có hộ gia đình nào được sưởi ấm bằng khí thiên nhiên.

b. Tìm xác suất mà có không nhiều hơn bốn trong số 20 hộ gia đình được sưởi ấm bằng khí thiên nhiên.

c. Tại sao thí nghiệm nhị thức có lẽ không cung cấp một mô hình tốt cho tình huống chọn mẫu này?

4.58 Giả sử, như đã được lưu ý trong Bài tập 4.19, rằng xấp xỉ khoảng 35% tất cả ứng viên tìm việc đã giả mạo thông tin trong hồ sơ xin việc của mình. Giả định một công ty có 2.300 nhân viên.

a. Giá trị kỳ vọng của con số x hồ sơ xin việc mà đã bị giả mạo thông tin là bao nhiêu?

b. Tìm độ lệch chuẩn của x.

c. Tính toán khoảng $(\mu \pm 2\sigma)$.

d. Giả định rằng hãng này có một công ty kiểm tra phẩm chất của người xin việc xác minh thông tin trên 2300 hồ sơ xin việc và rằng 249 hồ sơ xin việc có chứa đựng các thông tin giả mạo. Bạn có nghỉ rằng tỷ lệ giả mạo thông tin xin việc của công ty này là nhất quán với luận điểm rằng 35% tất cả các ứng viên xin việc giả mạo thông tin trên hồ sơ của mình không? Hãy giải thích.

4.59 Một kỹ sư kiểm tra chất lượng mong muốn nghiên cứu các phương án chọn mẫu thay thế $n = 5$, $a = 1$ và $n = 25$, $a = 5$. Trên cùng một tấm giấy kẻ để vẽ biểu đồ, hãy xây dựng các đường cong đặc trưng hoạt động cho cả hai phương án, bằng cách sử dụng các xác suất chấp nhận tại $p = 0{,}05$, $p = 0{,}10$, $p = 0{,}2$, $p = 0{,}3$ và $p = 0{,}4$ trong mỗi trường hợp.

a. Nếu bạn là một người bán sản xuất ra các lô hàng có tỷ lệ mắc lỗi trong khoảng từ $p = 0$ cho đến $p = 0{,}1$; thì phương án chọn mẫu nào trong số hai phương án kể trên sẽ được bạn ưa thích hơn?

b. Nếu bạn là một người mua mong muốn được bảo vệ để không phải chấp thuận tỷ lệ mắc lỗi vượt quá $p = 0{,}3$, thì phương án chọn mẫu nào trong số hai phương án kể trên sẽ được bạn ưa thích hơn?

4.60 Xem xét một phương án chấp nhận lô hàng với $n = 20$ và $a = 1$. Hãy tính toán xác suất của việc chấp thuận các lô hàng có các giá trị tỷ lệ bị mắc lỗi sau. Vẽ đường cong đặc trưng hoạt động cho phương án đó.

a $p = 0{,}01$

b $p = 0{,}05$

c $p = 0{,}1$

d $p = 0{,}2$

4.61 Một sự suy xét trong việc chọn lựa một nghề nghiệp là sự an toàn của người lao động trong ngành đó. Vào năm 1991, ngành công nghiệp sản xuất động cơ và thiết bị ô tô có tỷ lệ cao nhất về số thương tật nghề nghiệp, với 18,6 số ca thương tật trên mỗi 100.000 người lao động toàn thời gian, trong khi ngành nhà hàng (bao gồm các nơi ăn và uống) có tỷ lệ tai nạn thấp nhất, chỉ với 7,4 số vụ thương tật trên mỗi 100.000 người lao động toàn thời gian (*World Almanach and Book of Facts [Niên giám Thế giới và Sách Sự kiện],* phiên bản 1994). Giả sử rằng số vụ tai nạn trên mỗi 100.000 người lao động toàn thời gian trong ngành nhà hàng đi theo phân phối Poisson với trung bình bằng với μ = 7,4 trên mỗi 100.000 người lao động, rằng chuỗi nhà hàng trên toàn quốc có xấp xỉ 100.000 người lao động toàn thời gian, và rằng con số 7,4 là chính xác.

a. Sử dụng dữ liệu in ra từ Minitab để đánh giá xác suất mà chuỗi đó không có tai nạn nào.

b. Sử dụng dữ liệu in ra từ Minitab để đánh giá xác suất mà chuỗi đó có bảy hay ít hơn số vụ tai nạn.

c. Tính toán khoảng $\mu \pm 2\sigma$. Tim xác suất mà con số vụ tai nạn trên mỗi 100,000 người sẽ nằm trong khoảng này. Liệu xác suất này có nhất quán với Qui tắc Thực chứng?

POISSON VỚI TRUNG BÌNH = 7,400

K	P (X ≤ K)	K	P (X ≤ K)
0	0,006		
1	0,0051	11	0,9265
2	0,0219	12	0,9609
3	0,0632	13	0,9805
4	0,1395	14	0,9908
5	0,2526	15	0,9969
6	0,3920	16	0,9983
7	0,5393	17	0,9993
8	0,6757	18	0,9997
9	0,7877	19	0,9999
10	0,8707	20	1,0000

4.62 Trong một mẫu ngẫu nhiên gồm 20 thư ký chuyên nghiệp, 15 người ưa thích máy photo A hơn so với máy photo B. Nếu các máy photo này được mong muốn như nhau, thì xác suất mà một người sẽ chọn máy A so với B là 0,5. Xác suất mà con số *x*, ưa thích máy A hơn, trong mẫu gồm 20 người là bằng với 14 hay lớn hơn nếu như p = 0,5 là bao nhiêu?

4.63 Tham khảo lại Bài tập 4.62. Nhà sản xuất A thuê một công ty tiếp thị để chọn mẫu 500 thư ký chuyên nghiệp để quyết định máy nào trong số hai máy photo này được ưa thích hơn, A hay B.

a. Nếu p = 0,5 thì hãy tìm trung bình và độ lệch chuẩn của con số *x* trong mẫu này mà ưa thích máy photo A hơn.

b. Giả định rằng 280 thư ký chuyên nghiệp trong mẫu gồm 500 người ưa thích máy photo A hơn. Liệu kết quả mẫu này có khả năng xảy ra nếu trên thực tế hai máy photo này được mong muốn như nhau? Hãy giải thích.

4.64 Ngành công nghiệp hàng không Hoa Kỳ luôn luôn quan tâm hết mức đến thành tích an toàn của các hãng hàng không của mình. Tuy nhiên, vào năm 1989 thì số tai nạn chết người đã tăng lên 11, một sự gia tăng lớn qua các năm 1981-1992 (ngoại trừ năm 1989), mà trong đó con số trung bình của các vụ tai nạn chết người chỉ có 3,5 mỗi năm (*Nguồn:* Ban An toàn Vận chuyển Quốc gia, *Niên giám Thế giới và Sách Sự kiện,* phiên bản 1993, trang 274). Giả định rằng con số các tai nạn hàng không chết người mỗi năm có thể được ước lượng xấp xỉ bởi biến số ngẫu nhiên Poisson với trung bình bằng với 3,5.

a. Xác suất của việc quan sát 11 hay nhiều hơn các tai nạn chết người trong năm 1989 nếu trên thực tế μ = 3,5 là bao nhiêu?

b. Biết rằng 11 tai nạn thật sự đã xảy ra vào năm 1989, liệu có khả năng rằng con số trung bình các vụ tai nạn chết người vẫn là 3,5 hay không? Hãy giải thích.

4.65 Sự bùng nổ của các vũ khí hạt nhân là một mối quan ngại hàng đầu của tất cả người Mỹ. Trên thực tế, gần 70% toàn bộ người Mỹ nghĩ rằng mục tiêu chính sách ngoại giao hàng đầu của Hoa Kỳ nên ngăn chặn sự lan tràn của vũ khí hạt nhân (*US News & World Report [Thời báo Tin tức Hoa Kỳ và Báo cáo Thế giới]*, ngày 23 tháng Năm, 1994). Để kiểm tra tính chính xác của lời khẳng định này, bạn chọn mẫu ngẫu nhiên 600 người Mỹ.

a. Con số kỳ vọng và độ lệch chuẩn của *x*, số người trong mẫu mà cho rằng mục tiêu hàng đầu của Hoa Kỳ nên là ngăn chặn sự lan tràn của vũ khí hạt nhân là bao nhiêu?

b. Giả sử rằng trong mẫu của bạn có 365 người trả lời mà cảm thấy rằng mục tiêu hàng đầu của Hoa Kỳ nên là ngăn chặn sự lan tràn của vũ khí hạt nhân. Liệu đây có phải là một quan sát bất thường, khi giả định rằng con số 70% là chính xác? [*Gợi ý:* Tính một điểm số z cho quan sát này.]

c. Dựa vào các kết quả của câu (b), thì bạn có thể rút ra những kết luận nào?

4.66 Một kiểm toán viên được chứng nhận (CPA) đã tìm thấy chín trong mười báo cáo kiểm toán công ty có chứa những sai sót trầm trọng. Kiểm toán viên này bắt đầu với mười tài khoản công ty đầu tiên để kiểm toán.

a. Xác suất mà kiểm toán viên đó tìm thấy chín tài khoản công ty có những sai sót nghiêm trọng là bao nhiêu?

b. Xác suất mà kiểm toán viên đó tìm thấy tối đa có chín tài khoản công ty có những sai sót nghiêm trọng là bao nhiêu?

c. Xác suất mà kiểm toán viên đó tìm thấy có ít nhất chín tài khoản công ty có những sai sót nghiêm trọng là bao nhiêu?

***4.67** Một nhà bán lẻ nhận được một chuyến hàng giao gồm 200 tivi xách tay. Để tự bảo vệ mình không phải nhận một chuyến hàng "kém", anh ta sẽ kiểm tra năm tivi và chấp thuận toàn bộ lô hàng nếu như anh ta quan sát thấy không có tivi nào bị lỗi. Giả sử thực sự có 20 chiếc tivi bị lỗi trong lô hàng giao.

a. Xác suất mà anh ta chấp thuận toàn bộ lô hàng giao là bao nhiêu?

b. Biết rằng nhà bán lẻ chấp thuận toàn bộ lô hàng, thì xác suất mà anh ta quan sát thấy đúng một chiếc tivi nào bị lỗi là bao nhiêu?

4.68 Một nghiên cứu do một ngân hàng thực hiện đã tìm thấy rằng con số trung bình của những sai sót giao dịch trên mỗi nhân viên thu ngân mỗi ngày là 1,5. Liệu có khả năng xảy ra rằng bất cứ một nhân viên thu ngân nào sẽ thực hiện nhiều hơn bốn sai sót giao dịch mỗi ngày không? Giải thích.

***4.69** Phân phối xác suất nhị thức có được cái tên của mình từ sự kiện rằng các giá trị của $p(x)$ cho n lần thử và xác suất của thành công p tương ứng với số hạng của sự khai triển của $(q+p)^n$. Ví dụ, với n = 2 lần thử,

$$(q+p)^2 = q^2 + 2pq + p^2$$

trong đó $p(0) = q^2$, $p(1) = 2pq$, và $p(2) = p^2$. Khai triển $(q+p)^3$. Sau đó sử dụng công thức này cho $p(x)$ để chứng minh rằng $p(0) = q^3$, $p(1) = 3q^2 p$, và vân vân.

***4.70** Bài tập này sẽ cho bạn một cơ hội để xem cách thức mà công thức cho một phân phối xác suất nhị thức được tạo ra.

a. Liệt kê các sự kiện đơn giản cho một thí nghiệm nhị thức với n = 3 và xác suất của thành công trong một lần thử duy nhất bằng với *p*.

b. Tính toán các xác suất đi cùng với những sự kiện giản đơn này.

c. Chỉ định các sự kiện đơn giản phù hợp cho các sự kiện $x = 0, 1, 2$, và 3.

d. Tính toán $p(x)$ cho $x = 0, 1, 2$, và 3 bằng cách giả định rằng các xác suất của những sự kiện đơn giản phù hợp.

e. Sử dụng công thức này cho $p(x)$, $n = 3$, để tìm ra $p(x)$ cho $x = 0, 1, 2$, và 3. So sánh các kết quả này với câu trả lời của bạn trong câu (d).

4.71 Trong một nỗ lực nhằm giảm thiểu con số những lời than phiền của những người mua bàn để đầu giường bị đổ sập do thiếu ốc vít và chốt, người ta đã quyết định sử dụng một phương án chấp nhận với $n = 10$ đơn vị sản phẩm mỗi ca được kiểm tra về ốc vít và chốt bị thiếu, với một con số chấp thuận $a = 0$.

a. Tìm các xác suất chấp nhận cho phương án này với $p = 0$; 0,05; 0,10; 0,20, và 0,30.

b. Sử dụng các kết quả của câu (a) để xây dựng một đường cong đặc trưng hoạt động.

c. Nếu một hay nhiều hơn các đơn vị sản phẩm được tìm thấy là có thiếu ốc vít và chốt, thì tất cả các đơn vị sản phẩm đối với một ca đã biết sẽ bị kiểm tra và các vật phẩm bị thiếu sẽ được thay thế. Nếu bạn là nhà sản xuất, bạn sẽ kỳ vọng phương án này vận hành tốt như thế nào?