Math 631: An Introduction to Stochastic Processes (S09)

Instructor: Wenbo Li (Office: 517b Ewing, Ph: 831-1865, Email: wli@math.udel.edu).

Office Hours: 12:30-2:00 Tu, 2:00-3:30 Th, or by appointment.

Textbook: *Introduction to Probability Models, Seventh/Eighth/Ninth Edition* by Sheldon Ross.

Course Outline: We will cover most of the book and additional selected topics depending on the interests of students.

Homework: Homework will be assigned weekly in class. Some of them will be discussed in class.

Tests: There will be two examinations during the semester and a take-home final examination. A one page summary is allowed for each examination.

Grading: 50 points for each examination.

This is a second course in Probability (prerequisite: math630 or equivalent) aimed at graduate students in the Applied Mathematics, Electrical Engineering, Computer Science, Statistics, Business and Economics, who expect their thesis work to involve probability.

This course will emphasize describing what's known and how to do calculations in a broader range of probability models. Students are encouraged to learn by doing exercises.

Additional Reference book: *Probability and Random Processes*, by G. Grimmett and D. Stirzaker, which covering most of the same topics as our textbook in more depth, and covering more topics,

The Real Reason Why Software Engineers Need Math

" ... The human brain finds it extremely difficult to cope with a new level of abstraction. This is why it was well into the 18th century before mathematicians became comfortable with zero and negative numbers, and why even today many people cannot accept the square root of minus-one as a genuine number.

But software engineering is all about abstraction. Every single concept, construct, and method is entirely abstract. Of course, it doesn't feel this way to most software engineers. But that's my point. The main benefit they got from the mathematics they learned in academia was the experience of rigorous reasoning with purely abstract objects and structures. ...

Though the payoff from learning (any) mathematics is greater for the computer professional than most other people, in today's society the benefits affect everyone. ..."

—Keith Devlin, Communications of the ACM, vol 44, no 10 (Oct 2001), pp. 21-2.

Review: Axioms of Probability

The sample space, denoted by Ω, is the set of all possible outcomes in an "experiments". An event is a subset of Ω.

A *probability measure* $\mathbb{P}$ on Ω is a function on subsets of Ω such that

(i). $0 \leq \mathbb{P}(E) \leq 1$ for any event $E \subset \Omega$;

(ii). $\mathbb{P}(\Omega) = 1$;

(iii). For mutually exclusive events $E_1, E_2, \cdots$, (i.e $E_i \cap E_j = \emptyset$, $i \neq j$), $\mathbb{P}\left(\cup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(E_i)$.

Basic properties of $\mathbb{P}$:

- $\mathbb{P}(\emptyset) = 0$.
- $\mathbb{P}(E^c) = 1 - \mathbb{P}(E)$.
- $\mathbb{P}(E \cup F) = \mathbb{P}(E) + \mathbb{P}(F) - \mathbb{P}(EF)$.

- **Inclusion-exclusion formula:**

$$\begin{aligned}\mathbb{P}(\cup_{i=1}^{n} E_i) &= \sum_{i=1}^{n} \mathbb{P}(E_i) - \sum_{i_1<i_2} \mathbb{P}(E_{i_1}E_{i_2}) + \cdots \\ &\quad + (-1)^{r+1} \sum_{i_1<i_2<\cdots<i_r} \mathbb{P}(E_{i_1}E_{i_2}\cdots E_{i_r}) \\ &\quad + \cdots + (-1)^{n+1}\mathbb{P}(E_1E_2\cdots E_n)\end{aligned}$$

where the summation

$$\sum_{i_1<i_2<\cdots<i_r} \mathbb{P}(E_{i_1}E_{i_2}\cdots E_{i_r})$$

is taken over all of the $\binom{n}{r}$ possible subsets of size r of the set $\{1, 2, \cdots n\}$.

Review: Conditional Probability

Definition: If $\mathbb{P}(F) > 0$, then

$$\mathbb{P}(E|F) = \frac{\mathbb{P}(EF)}{\mathbb{P}(F)}$$

- Reduced sample space
- The multiplication rule: $\mathbb{P}(EF) = \mathbb{P}(F) \cdot \mathbb{P}(E|F)$.
- $\mathbb{P}(\bullet|F)$ is a probability measure.

Review: Laws of Total Probability

Let F_1, $F_2, \cdots, F_n$ be a partition of the sample space Ω, i.e. $\cup_{i=1}^n F_i = \Omega$, $F_i F_j = \emptyset$ for $i \neq j$. Then

$$\mathbb{P}(E) = \sum_{i=1}^{n} \mathbb{P}(E|F_i) \cdot \mathbb{P}(F_i).$$

In particular,

$$\mathbb{P}(E) = \mathbb{P}(E|F) \cdot \mathbb{P}(F) + \mathbb{P}(E|F^c) \cdot \mathbb{P}(F^c)$$

and

$$\mathbb{P}(E|F) = \sum_{i=1}^{n} \mathbb{P}(E|F_i F) \cdot \mathbb{P}(F_i|F).$$

Bayes formula:

$$\mathbb{P}(F_j|E) = \frac{\mathbb{P}(EF_j)}{\mathbb{P}(E)} = \frac{\mathbb{P}(E|F_j)\mathbb{P}(F_j)}{\sum_{i=1}^n \mathbb{P}(E|F_i) \cdot \mathbb{P}(F_i)}$$

Review: Independent Events

Two events E and F are *independent* if

$$\mathbb{P}(E|F) = \mathbb{P}(E), \quad \text{or} \quad \mathbb{P}(EF) = \mathbb{P}(E) \cdot \mathbb{P}(F)$$

Def: The events $E_1, \cdots, E_n$ are independent if for every subset $\{i_1, \cdots, i_r\} \subset \{1, 2, \cdots, n\}$, $r \leq n$,

$$\mathbb{P}(E_{i_1} E_{i_2} \cdots E_{i_r}) = \mathbb{P}(E_{i_1})\mathbb{P}(E_{i_2}) \cdots \mathbb{P}(E_{i_r}).$$

And an infinite set of events are independent if every finite of them are independent.

• This is typical way in mathematics to extend a definition from two to finite many, and then to infinite many.

Review: Random Variables

Definition: A random variable (r.v.) is a real-valued function on the sample space Ω, i.e. $X : \Omega \to \mathbb{R}$.

The (cumulative) distribution function $F(x)$ or $F_X(x)$ of the random variable X is

$$(cdf) \qquad F(b) = \mathbb{P}(X \leq b) \quad -\infty < b < \infty$$

If X is discrete, then

$$\mathbb{E}\,g(X) = \sum_{\text{all}x} g(x)\mathbb{P}(X = x).$$

If X is continuous with p.d.f $f_X(x) = F_X{}'(x)$, then

$$\mathbb{E}\,g(X) = \int_{-\infty}^{\infty} g(x)f(x)dx.$$

The variance of a r.v. X is defined by

$$\text{Var}(X) = \mathbb{E}\,(X - \mu)^2, \quad \text{where} \quad \mu = \mathbb{E}\,X.$$

It is a measurement of the spread of X around its mean $\mu = \mathbb{E}\,X$. The standard deviation is $\text{SD}(X) = \sqrt{\text{Var}(X)}$.

Basic properties of Var(X):

- $\text{Var}(X) = \mathbb{E}\,X^2 - (\mathbb{E}\,X)^2$.
- For any $a, b \in \mathbb{R}$, $\text{Var}(aX + b) = a^2\text{Var}(X)$.
- For any $a \in \mathbb{R}$,

$$\mathbb{E}\,(X - a)^2 \geq \mathbb{E}\,(X - \mu)^2 = \text{Var}(X).$$

The Bernoulli and Binomial r.v's

The Bernoulli r.v. takes only two values 0 and 1 with

$$p(1) = \mathbb{P}(X = 1) = p, \quad p(0) = \mathbb{P}(X = 0) = 1 - p$$

where p is the probability of success. Note that $\mathbb{E}\,X = p$ and $\mathrm{Var}(X) = \mathbb{E}\,X^2 - (\mathbb{E}\,X)^2 = p(1-p)$.

Suppose that n independent trials, each of which results in a success with probability p and in a failure with probability $1-p$, are to be performed. If X represents the number of successes that occur in the n trials, then X is said to be a *binomial* random variable with parameters (n, p). we also write $X \sim \mathrm{bi}(n, p)$ to represent binomial r.v. The probability mass function is given by

$$p(i) = \mathbb{P}(X = i) = \binom{n}{i} p^i (1-p)^{n-i}, \quad 0 \le i \le n.$$

• The mean $\mathbb{E}\,X = np$ and the variance $\mathrm{Var}(X) = \mathbb{E}\,X^2 - (\mathbb{E}\,X)^2 = np(1-p)$.

The Geometric random variable

Suppose that independent trials each having a probability p, $0 < p < 1$, of being a success, are performed until a success occurs. Then the number of trials required, X, have the geometric distribution

$$p(n) = \mathbb{P}(X = n) = (1-p)^{n-1}p, \quad n = 1, 2, \cdots .$$

- The sum

$$\sum_{n=1}^{\infty} \mathbb{P}(X = n) = \sum_{n=1}^{\infty} (1-p)^{n-1}p = 1.$$

- The mean

$$\mathbb{E}\, X = \sum_{n=1}^{\infty} n \cdot \mathbb{P}(X = n) = \sum_{n=1}^{\infty} n(1-p)^{n-1}p = \frac{1}{p}.$$

- The variance

$$\text{Var}(X) = \mathbb{E}\, X^2 - (\mathbb{E}\, X)^2 = \frac{1-p}{p^2}.$$

- The memoryless property:

$$\mathbb{P}(X = m + n \,|\, X > m) = \mathbb{P}(X = n) \quad m, n \geq 1.$$

The negative binomial r.v.

Suppose that independent trials each having a probability p, $0 < p < 1$, of being a success, are performed until a total of r successes is accumulated. Then the number of trials required, X, have the negative binomial distribution

$$\mathbb{P}(X = n) = \binom{n-1}{r-1} p^r (1-p)^{n-r}, \quad n = r, r+1, \cdots .$$

- The sum

$$\sum_{n=1}^{\infty} \mathbb{P}(X = n) = \sum_{n=1}^{\infty} \binom{n-1}{r-1} p^r (1-p)^{n-r} = 1.$$

- The mean

$$\mathbb{E}\, X = \sum_{n=1}^{\infty} n \cdot \binom{n-1}{r-1} p^r (1-p)^{n-r} = \frac{r}{p}.$$

- The variance

$$\mathrm{Var}(X) = \mathbb{E}\, X^2 - (\mathbb{E}\, X)^2 = \frac{r(1-p)}{p^2}.$$

Revisit of the negative binomial r.v.

In connection with the negative multinomial distribution, $Y = X - r$, the number of failures before r successes, is sometime called the negative binomial r.v. with

$$\mathbb{P}(Y = k) = \binom{k+r-1}{r-1} p^r (1-p)^k, \quad k = 0, 1, 2, \cdots .$$

- The negative binomial r.v. Y can be represented as

$$Y = (Y_1 - 1) + (Y_2 - 1) + \cdots + (Y_r - 1)$$

where Y_1 equals the number of trials required for the first success, Y_2 the number of additional trials after the first success until the second success occurs, and Y_j the number of additional trials after the $(j-1)$th success until the jth success occurs, $1 \le j \le r$. Note that Y_j are ind. geometric r.v's. Thus

$$\mathbb{E}\,Y = \mathbb{E}\,(Y_1 + \cdots + Y_r - r) = \mathbb{E}\,Y_1 + \cdots + \mathbb{E}\,Y_r - r = r\left(\frac{1}{p} - 1\right)$$

and

$$\mathrm{Var}(Y) = \mathrm{Var}(Y_1 + \cdots + Y_r) = \mathrm{Var}(Y_1) + \cdots + \mathrm{Var}(Y_r) = \frac{r(1-p)}{p^2}.$$

Review: The Poisson random variable

A r.v. X, taking on one of the values $0, 1, 2, \cdots$, is said to be a *Poisson* r.v. with parameter λ if for some $\lambda > 0$,

$$p(i) = \mathbb{P}(X = i) = e^{-\lambda}\frac{\lambda^i}{i!}, \quad i = 0, 1, 2, \cdots.$$

- The sum

$$\sum_{i=0}^{\infty} \mathbb{P}(X = i) = e^{-\lambda} \sum_{i=0}^{\infty} \frac{\lambda^i}{i!} = 1.$$

- The mean

$$\mathbb{E}\, X = \sum_{i=0}^{\infty} i \cdot \mathbb{P}(X = i) = e^{-\lambda} \sum_{i=0}^{\infty} i \cdot \frac{\lambda^i}{i!} = \lambda.$$

- The variance

$$\mathrm{Var}(X) = \mathbb{E}\, X^2 - (\mathbb{E}\, X)^2 = (\lambda^2 + \lambda) - \lambda^2 = \lambda.$$

- **Poisson Approximation:** Let $X \sim \mathrm{bi}(n, p_n)$ and $\lambda = n \cdot p_n$ fixed. Then

$$\mathbb{P}(X = i) \approx e^{-\lambda}\frac{\lambda^i}{i!} \quad \text{for } n \text{ large.}$$

Note that p_n is small for n large.

The Uniform Random Variable

A r.v X is *uniform* on the interval (a, b) if its p.d.f is given by

$$f(x) = \begin{cases} \frac{1}{b-a} & a < x < b, \\ 0 & \text{otherwise} \end{cases}$$

- The expectation $\mathbb{E}\, X = \int_a^b x \cdot \frac{1}{b-a} dx = (a+b)/2$.
- The variance $\text{Var}(X) = (b-a)^2/12$.

Exponential Random Variable

A continuous r.v X is *exponential* with parameter $\lambda > 0$ if its p.d.f is given by

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0, \\ 0 & \text{otherwise} \end{cases}$$

- The expectation $\mathbb{E}\, X = \int_0^\infty x \lambda e^{-\lambda x} dx = 1/\lambda$.
- The variance $\text{Var}(X) = 1/\lambda^2$.
- The *memoryless* property: For all $s, t \geq 0$,

$$\mathbb{P}(X > s + t \,|\, X > t) = \mathbb{P}(X > s).$$

Gamma Random Variables

A continuous r.v $X = X(\alpha, \lambda)$ is *Gamma* with parameter $\lambda > 0$ and $\alpha > 0$ if its p.d.f is given by

$$f(x) = \begin{cases} \frac{(\lambda x)^{\alpha-1}}{\Gamma(\alpha)} \cdot \lambda e^{-\lambda x} & x \geq 0, \\ 0 & \text{otherwise} \end{cases}$$

where $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$ is the gamma function.

- $\Gamma(s+1) = s\Gamma(s)$ for all $s \geq 0$ and $\Gamma(n+1) = n!$.
- The expectation $\mathbb{E}\, X = \int_0^\infty x \frac{(\lambda x)^{\alpha-1}}{\Gamma(\alpha)} \cdot \lambda e^{-\lambda x} dx = \alpha/\lambda$.
- The variance $\text{Var}(X) = \alpha/\lambda^2$.
- The distribution identity

$$X(n, \lambda) =^d e_1(\lambda) + \cdots + e_n(\lambda)$$

where e_i are i.i.d exponential r.v's with parameter $\lambda > 0$.

Review: Normal Random Variables

A r.v X is *normal* with parameter μ and σ^2, $\sigma > 0$, if its p.d.f is given by

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2} \qquad -\infty < x < \infty.$$

- One can show that $\int_{-\infty}^{\infty} f(x)dx = 1$.
- The expectation $\mathbb{E}\, X = \mu$.
- The variance $\text{Var}(X) = \sigma^2$ and $\text{SD}(X) = \sigma$.
- We use $X \sim N(\mu, \sigma^2)$ to denote normal r.v. with mean μ and variance σ^2.
- When $\mu = 0$ and $\sigma = 1$, $N(0,1)$ is called standard normal and denoted by Z with c.d.f

$$\Phi(x) = \mathbb{P}(Z \le x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt$$

and $\Phi(-x) = 1 - \Phi(x)$.
- Standardization: If $X \sim N(\mu, \sigma^2)$, then

$$Z = \frac{X - \mu}{\sigma} \sim N(0,1).$$

Review: Joint Distribution Functions

Definition: For any two r.v.'s X and Y, the *joint cumulative probability distribution function* of X and Y is defined by

$$F_{X,Y}(a,b) = \mathbb{P}(X \le a,\, Y \le b), \quad -\infty < a, b < \infty$$

The distribution of X, also called marginal, can be obtained from the joint one, $F_X(a) = \mathbb{P}(X \le a) = F(a, \infty)$. If both X and Y are discrete, then joint probability mass function is defined by $p_{X,Y}(x,y) = \mathbb{P}(X = x,\, Y = y)$. Thus

$$p_X(x) = \mathbb{P}(X = x) = \sum_y p_{X,Y}(x,y).$$

Jointly Continuous Random Variables

The non-negative function $f(x,y)$ is called *joint probability density function* (j.p.d.f) of X and Y if for any nice/measurable set $C \subset \mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$

$$\mathbb{P}((X,Y) \in C) = \int\int_C f(x,y)\,dx\,dy.$$

- Taking $C = \mathbb{R}^2$, then $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy = 1$.
- Taking $C = \{(x,y) : x \in A,\, y \in B\}$, then

$$\mathbb{P}(X \in A,\, Y \in B) = \int_A \int_B f(x,y)dydx.$$

- Connection between j.c.d.f. and j.p.d.f.

$$F_{X,Y}(a,b) = \mathbb{P}(X \le a,\, Y \le b) = \int_{-\infty}^{a}\int_{-\infty}^{b} f(x,y)dydx,$$

and $f(a,b) = \dfrac{\partial^2}{\partial a \partial b} F(a,b)$.

- Marginal p.d.f of X:

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy, \quad f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx.$$

- It is useful to think

$$f(x,y)dxdy \cong \mathbb{P}(x \le X \le x + dx,\, y \le Y \le y + dy) \cong \mathbb{P}(X \approx x,\, Y \approx y).$$

- $\mathrm{Cov}(X,Y) = \mathbb{E}\,(X - \mathbb{E}\,X)(Y - \mathbb{E}\,Y) = \mathbb{E}\,(XY) - \mathbb{E}\,X \cdot \mathbb{E}\,Y$ and

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

Review: Independent Random Variables

The r.v's X and Y are *independent* if for any two nice/measureable sets of real numbers A and B,

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B).$$

i.e., X and Y are independent if for all nice/measureable A and B, the events $E_A = \{X \in A\}$ and $F_B = \{Y \in B\}$ are independent.

- X and Y are ind. iff for all $a, b \in \mathbb{R}$,

$$\mathbb{P}(X \leq a, Y \leq b) = \mathbb{P}(X \leq a)\mathbb{P}(Y \leq b)$$

i.e. $F_{X,Y}(a, b) = F_X(a)F_Y(b)$ for all $a, b \in \mathbb{R}$.

- X and Y are ind. iff for all $x, y \in \mathbb{R}$,

$$f_{X,Y}(x, y) = f_X(x)f_Y(y) \quad \text{or} \quad p(x, y) = p_X(x)p_Y(y).$$

- X and Y are ind. iff their joint p.d.f can be expressed as

$$f_{X,Y}(x, y) = h(x)g(y), \quad -\infty < x, y < \infty.$$

- If X and Y are independent r.v's, then

$$\mathbb{E}\left(g(X)h(Y)\right) = \mathbb{E}\left(g(X)\right)\mathbb{E}\left(h(Y)\right).$$

To be covered in later sections

- Multinomial (MN)
- Negative multinomial (NMN)
- Connections and distribution representations.
- NMN from NMN and MN
- NMN from Poisson of Gamma
- see my notes.

Review: Conditional Distributions

The *conditional* probability mass function of X given $Y = y$ is

$$p_{X|Y}(x|y) = \mathbb{P}(X = x \,|\, Y = y) = \frac{p(x, y)}{p_Y(y)}$$

The *conditional* probability density function of X given $Y = y$ is

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

and the conditional c.d.f is

$$F_{X|Y}(a|y) = \mathbb{P}(X \leq a \,|\, Y = y) = \int_{-\infty}^{a} f_{X|Y}(x|y)dx.$$

- If X and Y are ind., then $f_{X|Y}(x|y) = f_X(x)$.
- Geometric meaning.
- $f_{X|Y}(x|y)dx \approx \mathbb{P}(x \leq X \leq x + dx|y \leq Y \leq y + dy)$.

Conditional Expectation

The *conditional expectation* of X given $Y = y$ is

$$\begin{aligned}\mathbb{E}(X|Y=y) &= \textstyle\sum_x x \cdot \mathbb{P}(X=x|Y=y) \\ &\left(= \textstyle\int_{-\infty}^{\infty} x \cdot f_{X|Y}(x|y)dx\right)\end{aligned}$$

Note that $\mathbb{E}(X|Y=y)$ is a function of y.

Computing expectations by conditioning

Let $\mathbb{E}(X|Y)$ be a function of r.v. Y whose value at $Y = y$ is $\mathbb{E}(X|Y=y)$. Then

$$\mathbb{E}X = \mathbb{E}(\mathbb{E}(X|Y))$$

If Y is discrete (resp. conti.), then the above states

$$\mathbb{E}X = \sum_y \mathbb{E}(X|Y=y) \cdot \mathbb{P}(Y=y)$$

$$(\text{resp.} \quad \mathbb{E}X = \int_{-\infty}^{\infty} \mathbb{E}(X|Y=y) \cdot f_Y(y)dy$$

Ex (The expectation of the sum of random numbers of r.v's): Let X_i be i.i.d r.v's with $\mathbb{E}\,X_i = \mu$. Assume the positive integer-valued r.v N is independent of X_i, $i \geq 1$, and $\mathbb{E}\,N < \infty$. Then

$$\mathbb{E}\left(\sum_{i=1}^{N} X_i\right) = \cdots$$

Ex (The mean of a Geometric distribution): Suppose that independent trials each having a probability p, $0 < p < 1$, of being a success, are performed until a success occurs. What is the expected number of trials required?

Sol 1. Let N be the number of trials required. Then N have the geometric distribution $p(n) = \mathbb{P}(N = n) = (1-p)^{n-1}p, n = 1, 2, \cdots$ and hence

$$\mathbb{E}\,N = \sum_{n=1}^{\infty} n \cdot \mathbb{P}(N = n) = \sum_{n=1}^{\infty} n(1-p)^{n-1}p = \frac{1}{p}.$$

Sol 2. Let $Y = 1$ if the first trial is a success, and $Y = 0$ otherwise. Then

$$\mathbb{E}\,N = \mathbb{E}\,(N|Y = 1) \cdot \mathbb{P}(Y = 1) + \mathbb{E}\,(N|Y = 0) \cdot \mathbb{P}(Y = 0) = \cdots$$

- Which is better?

The matching problem. (Example 3.22, 2.31, 3.13 and Exercise 2.66-68).
Suppose that each of n men at a party throws his hat into the center of the room. The hats are first mixed up, and then each man randomly selects a hat.

(a) Find the probability that none of the men selects his own hat.
Sol. I (Inclusion-exclusion formula).
Sol. II (First Step Analysis).
Sol. III (Conditioning on the Cycle Size).

(b). Find the probability that exactly k of the men select their own hats.
Ans: Let X_n be the number of matches with n men. Then

$$\mathbb{P}(X_n = k) = \frac{1}{k!}\sum_{i=0}^{n-k}(-1)^i\frac{1}{i!}.$$

(c). Find the expected number of people that select their own hats.
Sol. 1. "Hard" algebraic method.

$$\mathbb{E}\,X_n = \sum_{k=0}^{n} k \cdot \mathbb{P}(X_n = k) = \cdots = 1.$$

Sol. 2. Decomposition as a sum. Namely $X_n = \sum_{i=1}^{n} Y_i$ where $Y_i = 1$ if the i^{th} man selects his own hat, and $Y_i = 0$ otherwise. Then $\mathbb{E}\, X_n = \sum_{i=1}^{n} \mathbb{E}\, Y_i = 1$ since $\mathbb{E}\, Y_i = \mathbb{P}(Y_i = 1) = 1/n$.

(d). Find the variance of the number of people that select their own hats.
Sol. 1. Very "Hard" algebraic method. $\text{Var}(X_n) = \mathbb{E}\, X_n^2 - (\mathbb{E}\, X_n)^2$ and

$$\mathbb{E}\, X_n^2 = \sum_{k=0}^{n} k^2 \cdot \mathbb{P}(X_n = k) = \cdots = 2$$

Sol. 2. Decomposition as a sum. Use

$$\text{Var}(X_n) = \sum_{i=1}^{n} \text{Var}(Y_i) + 2 \sum_{i<j} \text{Cov}(Y_i, Y_j)$$

and

$$\begin{aligned}
\text{Var}(Y_i) &= \mathbb{E}\, Y_i^2 - (\mathbb{E}\, Y_i)^2 = (1/n) - (1/n)^2 \\
\text{Cov}(Y_i, Y_j) &= \mathbb{E}\,(Y_i Y_j) - \mathbb{E}\, Y_i \cdot \mathbb{E}\, Y_j \\
\mathbb{E}\,(Y_i Y_j) &= \mathbb{P}(Y_i = 1, Y_j = 1) = \frac{1}{n} \cdot \frac{1}{n-1}.
\end{aligned}$$

We have $\text{Var}(X_n) = 1$.
- X_n can be approximated by Poisson r.v. with $\lambda = 1$.

Now suppose that those choosing their own hats depart, while the others (those without a match) put their selected hats in the center of the room, mixed them up, and then reselect. Also, suppose that this process continues until each individual has his own hat.

(e). Find $\mathbb{E}\, R_n$ where R_n is the number of rounds that are necessary.

(f). Find $\mathbb{E}\, S_n$ where S_n is the total number of selections.

(g). Find the expected number of false selections made by one of the n man.

(h)*. Find $\mathbb{E}\, L_n$ where L_n is the number of men on the last round.

Ref: W.V. Li, F. Liu and X. Shi (2006), An analysis of the last round matching problem, *Journal of Mathematical Analysis and Applications*, **323**, 1373–1382.

Ex 3.14: Independent trials, each of which is a success with prob. p, are performed until there are k consecutive successes. What is the mean number of necessary trials?

Sol 1: Use the first step analysis. It is good for small values of k and suggests the conditioning in Sol. 2.
Sol 2: Conditioning on what matters (can be dealt with) and looking for recursive relations.

Let N_k denote the number of necessary trials to obtain k consecutive successes (what is asked), and let $M_k = \mathbb{E}\, N_k$. Then

$$M_k = \mathbb{E}\, N_k = \mathbb{E}\,(\mathbb{E}\,(N_k|N_{k-1}))$$

and

$$\mathbb{E}\,(N_k|N_{k-1}) = p(N_{k-1} + 1) + (1-p)(N_{k-1} + 1 + \mathbb{E}\, N_k).$$

Thus

$$M_k = \frac{1}{p} + \frac{M_{k-1}}{p} \quad \text{and} \quad M_k = \frac{1}{p} + \frac{1}{p^2} + \cdots + \frac{1}{p^k}.$$

Sequential Sampling Without Replacement

Assume that we have $m+k$ balls in an urn and m of them are red (all others are black). We sequentially sample (select at random one ball at a time) from the urn *without* replacement. What is the expected number of tries to obtain the first red ball?

Sol: There are many different proofs. One of the simplest is by the recursive relation from the first step conditioning

$$\begin{aligned} E(m,k) = \mathbb{E}\, N_{m,k} &= \mathbb{E}\,(N_{m,k}|R) \cdot \frac{m}{m+k} + \mathbb{E}\,(N_{m,k}|R^c) \cdot \frac{k}{m+k} \\ &= 1 \cdot \frac{m}{m+k} + (1 + E(m,k-1)) \cdot \frac{k}{m+k}. \end{aligned}$$

•Can you guess the answer? Is it right? Any difference with the sampling *with* replacement? Which is smaller?

•What is the expected number of tries to obtain the last red ball?

$$L(m,k) = (1 + L(m-1,k)) \cdot \frac{m}{m+k} + (1 + E(m,k-1)) \cdot \frac{k}{m+k}.$$

and hence $L(m,k) = m(m+k+1)/(m+1)$ by the guess-check method.

An Alternative Sol: Representation as Sums

EX: (From Math 630) A box contain 9 light bulbs, of which 2 are defective. What is the expected value of the number of light bulbs that one will have to test (at random and without replacement) to find both defective bulbs?

Sol: Let $X_{ij} = j$ if the i-th and j-th light bulbs to be examined are defective, and $X_{ij} = 0$ otherwise, for $1 \le i < j \le 9$. Then $\sum_{1 \le i < j \le 9} X_{ij}$ is the number of light bulbs to be examined. Note that $\mathbb{P}(X_{ij} = j) = 1/\binom{9}{2}$.

•In the general setting, we can define $X_{i_1 \cdots i_m} = i_m$ if the selection on $i_1, \cdots, i_m$ are red balls, for $1 \le i_1 < \cdots < i_m \le m + k$, and zero otherwise. Thus

$$\begin{aligned} L(m,k) &= \mathbb{E} \sum_{1 \le i_1 < \cdots < i_m \le m+k} X_{i_1 \cdots i_m} = \sum i_m \cdot \mathbb{P}(X_{i_1 \cdots i_m} = i_m) \\ &= \binom{m+k}{m}^{-1} \sum_{m \le i_m \le m+k} i_m \cdot \sum_{1 \le i_1 < \cdots < i_m} 1 = \ldots\ldots \\ &= m(m+k+1)/(m+1) \end{aligned}$$

•Which method is better: Conditioning or Representation?

Ex 3.15: Analyzing the Quick-Sort Algorithm. (The First Step Analysis). Suppose we are given a set of n distinct values–$x_1, \cdots, x_n$–and we desire to put these values in increasing order or *sort* them. An efficient procedure for accomplishing this is the quick-sort algorithm which is defined recursively as follows: It starts by choosing at random one of the n values–say, x_i–and then compares each of the other $n-1$ values with x_i, noting which are smaller and which are larger than x_i. Letting S_i denote the set of elements smaller than x_i, and $\bar{S}_i$ the set of elements greater than x_i, the algorithm now sorts the set S_i and the set $\bar{S}_i$. The final ordering, therefore, consists of the ordered set of the elements in S_i, then x_i, and then the ordered set of the elements in $\bar{S}_i$.

To measure the effectiveness of this algorithm, let N_n be the number of comparisons. Then by simple conditioning argument,

$$\mathbb{E}\, N_n = \sum_{j=1}^{n} (n-1+\mathbb{E}\, N_{j-1}+\mathbb{E}\, N_{n-j})\frac{1}{n} = n-1+\frac{2}{n}\sum_{k=1}^{n-1}\mathbb{E}\, N_k$$

which implies $\mathbb{E}\, N_n \sim 2n \ln n$.

• The asymptotic range of N_n is $(n\log_2 n, n^2/2)$ and $\mathrm{SD}(N_n) = \sqrt{\mathrm{Var}(N_n)} \sim n\sqrt{7-\frac{2}{3}\pi^2}$. The normalized variate $(N_n - \mathbb{E}\, N_n)/n \Longrightarrow Y$ where Y is a well-defined r.v with unknown density.

3.6.1. A List Model: Consider n elements–$e_1, e_2, \cdots, e_n$–that are initially arranged in some ordered list. At each unit of time a request is made for one of these elements–e_i being requested, independent of the past, with probability P_i. After being requested the element is then moved to the front of the list. What is the expected position of the element requested after this process has been in operation for a long time?

• This list model will be compared with the one closer model in Sec. 4.8.

Sol: (Conditioning, asymptotic independents, long run probabilities, representation as sums, etc).

$$\begin{aligned}
& \mathbb{E}\,(\text{position of element requested}) \\
= & \sum_{i=1}^{n} \mathbb{E}\,(\text{position}|e_i \text{ is selected}) \cdot P_i \\
= & \sum_{i=1}^{n} \mathbb{E}\,(\text{position of } e_i|e_i \text{ is selected}) \cdot P_i \\
= & \sum_{i=1}^{n} \mathbb{E}\,(\text{position of } e_i) \cdot P_i
\end{aligned}$$

and note that

$$\text{position of } e_i = 1 + \sum_{j \neq i} I_{ji},$$

where $I_{ji} = \begin{cases} 1 & \text{if } e_j \text{ precedes } e_i \\ 0 & \text{otherwise.} \end{cases}$ Thus

$$\begin{aligned} \mathbb{E}\,(\text{position of } e_i) &= 1 + \sum_{j \neq i} \mathbb{E}\, I_{ji} \\ &= 1 + \sum_{j \neq i} \mathbb{P}(e_j \text{ precedes } e_i) \\ &= 1 + \sum_{j \neq i} \frac{P_j}{P_i + P_j} \end{aligned}$$

and the expected position of element requested

$$\mathbb{E}\,(\text{position}) = 1 + \sum_{i=1}^{n} P_i \sum_{j \neq i} \frac{P_j}{P_i + P_j}.$$

Homework

Ch3: 7th edition (8th) [9th].

2*. Let X_1 and X_2 be independent geometric random variables having the same parameter p. Guess and find the value of $\mathbb{P}(X_1 = i \,|\, X_1 + X_2 = n)$.

8. An unbiased die is successively rolled. Let X and Y denote respectively the number of rolls necessary to obtain a six and a five. Find (a). $\mathbb{E}\,X$; (b). $\mathbb{E}\,(X|Y = 1)$; (c). $\mathbb{E}\,(X|Y = 5)$.

13*. Let X be exponential with mean $1/\lambda$. Find $\mathbb{E}\,(X|X > 1)$.

23*, **24**, **26**, **27**, **53**.

Ch4: 7th edition (8th) [9th]. **1***, **2**, **4***, **12*** [**16***], **20** [**24**], **21**[**25**], **22**[**26**], **23***[**27***], **26**[**30**], **28***[**32***], **32***[**41***], **33**[**42**], **34**[**43**], **38***[**47***], **40**[**49**], **44**[**55**], **47**[**58**], **51***[**62***], **56**[**67**], **57***[**68***], **59**[**70**], **63**[**74**].