Math Review Notes

Gregory Faletto

Contents

Last updated November 27, 2018

1 Linear Algebra

1.1 Properties of Projection Matrices

i. Formula:

$$P = A(A^T A)^{-1} A^T$$

(Note that if A is an invertible (square) matrix, then $P = A(A^T A)^{-1} A^T = AA^{-1}(A^T)^{-1} A^T = I$.)

The projection matrix projects any vector b into the column space of A. In other words, $p = Pb$ is the component of b in the column space, and the error $e = b - Pb$ is the component in the orthogonal complement. ($I - P$ is also a projection matrix. It projects b onto the orthogonal complement, and the projection is $b - Pb = e$).

(Note that if A is an invertible (square) matrix, then its column space is all of $\mathbb{R}^n$, so b is already in the column space of A.)

ii. The projection matrix is **idempotent**: it equals its square–$P^2 = P$.

iii. The projection matrix is **symmetric**: it equals its transpose–$P^T = P$.

iv. Conversely, **any symmetric idempotent matrix represents a projection**. P is unique for a given subspace.

v. If A is an $m \times n$ matrix with rank n, then $\text{rank}(P) = n$. The eigenvalues of P consist of n ones and $m - n$ zeroes. P always contains n independent eigenvectors and is thus diagonalizable.

Suppose A is a square nonsingular matrix and λ is an eigenvalue of A. Then λ^{-1} is an eigenvalue of the matrix A^{-1}.

The trace of an idempotent matrix with rank r is r.

1.2 Eigenvalues, Eigenvectors, Diagonalization, Symmetric Matrices

Notes on Diagonalization

Suppose the $n \times n$ matrix A has n linearly independent eigenvectors. If these eigenvectors are the columns of a matrix S, then $S^{-1}AS$ is a diagonal matrix Λ. The eigenvalues of A are on the diagonal of Λ:

$$S^{-1}AS = \Lambda = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}$$

We call S the **eigenvector matrix** and Λ the **eigenvalue matrix**.

1. If the matrix A has no repeated eigenvalues, then its n eigenvectors are automatically independent. Therefore **any matrix with n distinct eigenvalues can be diagonalized**.

2. **The diagonalizing matrix S is not unique**. An eigenvector x can be multiplied by a constant and remains an eigenvector. We can multiply the columns of S by any nonzero constants and produce a new diagonalizing S. Repeated eigenvalues leave even more freedom in S (columns with identical eigenvalues can be interchanged).

   (Note that for the trivial example $A = I$, any invertible S will do. $S^{-1}IS$ is always diagonal, and Λ is just I. **All vectors are eigenvectors of the identity.**)

3. **Other matrices S will not produce a diagonal Λ**. Since $\Lambda = S^{-1}AS$, S must satisfy $S\Lambda = AS$. Suppose the first column of S is y. Then the first column of $S\Lambda$ is $\lambda_1 y$. If this is to agree with the first column of AS, which by matrix multiplication is Ay, then y must be an eigenvector: $Ay = \lambda_1 y$.

   (Note that the *order* of the eigenvectors in S and the eigenvalues in Λ must match.)

4. Not all matrices posses n linearly independent eigenvectors, so **not all matrices are diagonalizable**.

   Diagonalizability of A depends on having enough (n) independent eigenvectors. Invertibility of A depends on having nonzero eigenvalues.

   There is no connection between diagonalizability (n independent eigenvectors) and invertibility (no zero eigenvalues). The only indication given by the eigenvalues is that diagonalization can fail only if there are repeated eigenvalues. (But even then, it does not always fail–e.g. I.)

   The test is to check, for an eigenvalue that is repeated p times, whether there are p independent eigenvectors–in other words, whether $A - \lambda$ has rank $n - p$.

5. **Projection matrices always contain n independent eigenvectors and thus are always diagonalizable**.

Eigenvalues of Symmetric Matrices: If A is symmetric, then it has the following properties:

1. A has exactly n (not necessarily distinct) eigenvalues
2. There exists a set of n eigenvectors, one for each eigenvalue, that are mutually orthogonal (even if the eigenvalues are not distinct).

Eigenvalues of the Inverse of a Matrix: Suppose A is a square nonsingular matrix and λ is an eigenvalue of A. Then λ^{-1} is an eigenvalue of the matrix A^{-1}. Proof: Note that since A is nonsingular, A^{-1} exists and λ is nonnegative for all eigenvalues of A. Let λ be an eigenvalue of A and let $x \neq 0$ be an eigenvector of A for λ. Suppose A is n by n. Then we have

$$A^{-1}x = A^{-1}\lambda^{-1}\lambda x = \lambda^{-1}A^{-1}\lambda x = \lambda^{-1}A^{-1}Ax = \lambda^{-1}x$$

The inverse of a symmetric matrix is symmetric. Proof: Let A be a symmetric matrix.

$$I = I'$$

$$AA^{-1} = (AA^{-1})'$$

$$A^{-1}A = (A^{-1})'A'$$

$$A^{-1}AA^{-1} = (A^{-1})'AA^{-1}$$

$$A^{-1} = (A^{-1})'$$

1.3 Positive Definite Matrices

For any real invertible matrix A, the product $A'A$ is a positive definite matrix. (Proof: Let z be a non-zero vector. We want $z'A'Az > 0 \forall z$. Note that $z'A'Az = (Az)'(Az)$. Because A is invertible and $z \neq 0$, $Az \neq 0$, so $(Az)'(Az) > 0$.)

Let $A \in \mathbb{R}^{m \times n}$ with $m \geq n$ and let $\text{rank}(A) = n$ (that is, A has full column rank). Then $A'A$ is a positive definite matrix. (Proof: Let z be a non-zero vector. We want $z'A'Az > 0 \forall z$. Note that $z'A'Az = (Az)'(Az)$. Because A has full column rank (and n linearly independent columns) and $z \neq 0$, $Az \neq 0$, so $(Az)'(Az) > 0$.)

Every positive definite matrix is invertible and its inverse is also positive definite.

1.4 Other

Frobenius norm

QR decompositon

From appendix of Convex Optimization:

Orthogonal Decomposition

Spectral Decomposition (eigenvalue decomposition

Generalized eigenvalue decomposition

Singular value decomposition

Pseudo-inverse (Moore-Penrose inverse)

Schur complement, Schur decomposition

From appendix of Time Series:

Quadratic forms

Special matrices

Jordan decomposition

Cholesky decomposition

Difference Equations

1.5 Practice Problems

12. Let A be a 2×2 matrix for which there is a constant k such that the sum of the entries in each row and each column is k. Which of the following must be an eigenvector of A ?

I. $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$

II. $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$

III. $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$

(A) I only (B) II only (C) III only (D) I and II only (E) I, II, and III

Solution 12. (C) This condition makes the matrix of the form

$$\begin{pmatrix} a & b \\ b & a \end{pmatrix}.$$

There is no reason that $a = 0$ or $b = 0$, so there is no reason $(1, 0)$ or $(0, 1)$ should be eigenvectors. But it is easy to verify that $(1, 1)$ must be.

24. Consider the system of linear equations

$$\begin{aligned} w + 3x + 2y + 2z &= 0 \\ w + 4x + y &= 0 \\ 3w + 5x + 10y + 14z &= 0 \\ 2w + 5x + 5y + 6z &= 0 \end{aligned}$$

with solutions of the form (w, x, y, z), where w, x, y, and z are real. Which of the following statements is FALSE?

(A) The system is consistent.

(B) The system has infinitely many solutions.

(C) The sum of any two solutions is a solution.

(D) $(-5, 1, 1, 0)$ is a solution.

(E) Every solution is a scalar multiple of $(-5, 1, 1, 0)$.

Solution 24. (E) Looking at our answers, we can verify directly that $(-5, 1, 1, 0)$ is a solution. Any multiple of $(-5, 1, 1, 0)$ is also a solution, which shows that (A), (B), (C), and (D) are all true – leaving only (E). Another solution, for example, is $(0, 2, -8, 5)$

$$A = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 0 & 2 & 3 & 4 & 5 \\ 0 & 0 & 3 & 4 & 5 \\ 0 & 0 & 0 & 4 & 5 \\ 0 & 0 & 0 & 0 & 5 \end{pmatrix}$$

34. Which of the following statements about the real matrix shown above is FALSE?

(A) A is invertible.

(B) If $\mathbf{x} \in \mathbb{R}^5$ and $A\mathbf{x} = \mathbf{x}$, then $\mathbf{x} = \mathbf{0}$.

(C) The last row of A^2 is $(0 \quad 0 \quad 0 \quad 0 \quad 25)$.

(D) A can be transformed into the 5×5 identity matrix by a sequence of elementary row operations.

(E) $\det(A) = 120$

Solution 34. (B) An upper triangular matrix is easily verified to be invertible so long as its diagonal entries are all nonzero. Specifically, $\det A$ is still the product of its diagonal entries, so (E) and (D) and (A) are all true. (C) can easily be verified to be true by computing that the bottom-right corner is 25 (the product of upper triangular matrices still being upper triangular). This leaves (B). (B) can be checked directly to be false: if we let $x = (1, 0, 0, 0, 0)$, then $Ax = x$.

37. Let V be a finite-dimensional real vector space and let P be a linear transformation of V such that $P^2 = P$. Which of the following must be true?

I. P is invertible.

II. P is diagonalizable.

III. P is either the identity transformation or the zero transformation.

(A) None (B) I only (C) II only (D) III only (E) II and III

Solution 37. (C) $P^2 = P$ means that P is projection onto some subspace. There is no reason to believe that this should be invertible, but it should definitely be diagonalisable (with eigenbasis some basis of that subspace). III also need not be true if the subspace is anything proper or nontrivial.

50. Let A be a real 2×2 matrix. Which of the following statements must be true?

I. All of the entries of A^2 are nonnegative.

II. The determinant of A^2 is nonnegative.

III. If A has two distinct eigenvalues, then A^2 has two distinct eigenvalues.

(A) I only (B) II only (C) III only (D) II and III only (E) I, II, and III

Solution 50. (B) There is no reason that all the entires of A^2 need to be nonnegative. Its determinant must be nonnegative though: $\det(A^2) = (\det A)^2$. For III, suppose A is the diagonal matrix with entires $\pm\lambda$. Then those are its eigenvalues, and they are distinct so long as $\lambda \neq 0$. But A^2 has only one eigenvalue: λ^2.

51. Which of the following is an <u>orthonormal</u> basis for the column space of the real matrix $\begin{pmatrix} 1 & -1 & 2 & -3 \\ -1 & 1 & -3 & 2 \\ 2 & -2 & 5 & -5 \end{pmatrix}$?

(A) $\left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \right\}$

(B) $\left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\}$

(C) $\left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{2}{\sqrt{5}} \\ \frac{1}{\sqrt{5}} \\ 0 \end{pmatrix} \right\}$

(D) $\left\{ \begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 \\ -3 \\ 5 \end{pmatrix} \right\}$

(E) $\left\{ \begin{pmatrix} \frac{1}{\sqrt{6}} \\ -\frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{6}} \end{pmatrix}, \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{pmatrix} \right\}$

Solution 51. (E) The basis (C) is not orthogonal and (D) is not normal, so we can rule those out. We can throw out the first column, since it is the negation of the second. A little bit of math shows that the remaining 3×3 matrix has determinant 0, so the rank of our column space is 2. That leaves only (A) and (E), but (A) cannot be correct. Our column space contains vectors that have nonzero third entry, so cannot lie in the span of that basis.

2 Calculus

These notes include some screenshots from Wikipedia as well as from *Calculus* by Gilbert Strang, available at https://ocw.mit.edu/ans7870/resources/Strang/Edited/Calculus/Calculus.pdf. I also used parts from some other resources which I mention when they arise.

2.1 List of common derivatives and integrals to know

$$\frac{d}{dx}\left(\sin^{-1} x\right) = \frac{1}{\sqrt{1-x^2}} \qquad \frac{d}{dx}\left(\ln(x)\right) = \frac{1}{x},\ x>0$$

$$\frac{d}{dx}\left(\cos^{-1} x\right) = -\frac{1}{\sqrt{1-x^2}} \qquad \frac{d}{dx}\left(\ln|x|\right) = \frac{1}{x},\ x\neq 0$$

$$\frac{d}{dx}\left(\tan^{-1} x\right) = \frac{1}{1+x^2} \qquad \frac{d}{dx}\left(\log_a(x)\right) = \frac{1}{x\ln a},\ x>0$$

$$\int \tan u\, du = \ln|\sec u| + c$$

$$\int \sec u\, du = \ln|\sec u + \tan u| + c$$

$$\int \frac{1}{a^2+u^2}\, du = \frac{1}{a}\tan^{-1}\left(\frac{u}{a}\right) + c$$

$$\int \frac{1}{\sqrt{a^2-u^2}}\, du = \sin^{-1}\left(\frac{u}{a}\right) + c$$

$$\int \ln u\, du = u\ln(u) - u + c$$

$$\int \sinh x\, dx = \cosh x + C \qquad \int \cosh x\, dx = \sinh x + C$$

2.2 Matrix Differentiation

Recommended resource: "Matrix Differentiation (and some other stuff)" by Randal J. Barnes (Department of Civil Engineering, University of Minnesota). Available for download at https://atmos.washington.edu/~dennis/MatrixCalculus.pdf.

More information not contained in that pdf (from the appendix of *Convex Optimization* by Stephen Boyd and Lieven Vandenberghe, available for free download at https://web.stanford.edu/~boyd/cvxbook/):

Chain rule. Suppose $f : \mathbb{R}^n \to \mathbb{R}^m$ is differentiable at $x \in \textbf{int dom} f$ and $g : \mathbb{R}^m \to \mathbb{R}^p$ is differentiable at $f(x) \in \textbf{int dom}\ g$. Define the composition $h : \mathbb{R}^n \to \mathbb{R}^p$ by $h(z) = g(f(x))$. Then

$$Dh(x) = Dg(f(x))Df(x)$$

In particular, if $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$,

$$\nabla h(x) = g'(f(x))\nabla f(x)$$

Example with an affine function. Suppose $f : \mathbb{R}^n \to \mathbb{R}^m$ is differentiable, $A \in \mathbb{R}^{n \times p}$, and $b \in \mathbb{R}^n$. Define $g : \mathbb{R}^p \to \mathbb{R}^m$ as $g(x) = f(Ax + b)$ with **dom** $g = \{x \mid Ax + b \in$ **dom** $f\}$. Then

$$\nabla g(x) = A^T \nabla f(Ax + b)$$

Example 2. Consider $f : \mathbb{R}^n \to \mathbb{R}$ where

$$f(x) = \log \sum_{i=1}^m \exp(a_i^T x + b_i) =$$

where $a_1, \ldots, a_m \in \mathbb{R}^n$ and $b_1, \ldots, b_m \in \mathbb{R}$. Note that $f(\cdot)$ can be expressed as a composition of $Ax + b$ (where $A \in \mathbb{R}^{m \times n}$ has rows $a_1^T, \ldots, a_m^T$) and the function $g : \mathbb{R}^m \to \mathbb{R}$ given by $g(y) = \log(\sum_{i=1}^m \exp(y_i)$. We have

$$\nabla g(y) = \left[\sum_{i=1}^m e^{y_i} \right]^{-1} \big(\exp(y_1) \ldots \exp(y_m) \big)^T$$

so applying the chain rule yields

$$\nabla f(x) = \left[\sum_{i=1}^m \exp(a_i^T x + b_i) \right]^{-1} A^T z$$

where $z_i = \exp(a_i^T x + b_i), \; i = 1, \ldots, m$.

Hessians. The Hessian matrix of $f : \mathbb{R}^n \to \mathbb{R}$ is denoted by $\nabla^2 f(x)$ and is given by

$$\nabla^2 f(x)_{ij} = \frac{\partial^2 f(x)}{\partial x_i \partial x_j}, \; i = 1, 2, \ldots, n, \quad j = 1, 2, \ldots, n$$

The quadratic function

$$f(x_0) + \nabla f(x_0)^T (x - x_0) + \frac{1}{2}(x - x_0)^T \nabla^2 f(x_0)(x - x_0)$$

is called the **second-order approximation of f near** x_0.

Chain rule for second derivative. A chain rule for the second derivative is difficult in general. Here are some special cases.

Composition with scalar function. Suppose $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R} \to \mathbb{R}$, and $h(x) = g(f(x))$. We have

$$\nabla^2 h(x) = g'(f(x))\nabla^2 f(x) + g''(f(x))\nabla f(x) \nabla f(x)^T$$

Composition with affine function. Suppose $f : \mathbb{R} \to \mathbb{R}$, $a \in \mathbb{R}^m$, $b \in \mathbb{R}$. Define $g : \mathbb{R}^m \to \mathbb{R}$ by $g(x) = f(a^T x + b)$. Then

$$\nabla^2 g(x) = a^T \nabla^2 f(a^T x + b) a$$

More generally, suppose $f : \mathbb{R}^n \to \mathbb{R}$, $A \in \mathbb{R}^{n\times m}$, and $b \in \mathbb{R}^n$. Define $g : \mathbb{R}^m \to \mathbb{R}$ by $g(x) = f(Ax + b)$. Then

$$\nabla^2 g(x) = A^T \nabla^2 f(Ax + b) A$$

2.3 Some theorems in higher dimensions

Taylor's Theorem (first order). (borrowed from https://www.rose-hulman.edu/~bryan/lottamath/mtaylor.pdf) Consider a function $f : \mathbb{R}^n \to \mathbb{R}$. Let $a \in \mathbb{R}^n$ be a fixed point. Then Taylor's Theorem states:

If $f(x)$ is differentiable on an open ball B around a and $x \in B$ then

$$f(x) = f(a) + \nabla f(b)^T (x - a)$$

for some b on the line segment joining a and x.

This can also be expressed as follows. *Let $x, y \in \mathbb{R}^n$. If $f(x)$ is continuously differentiable, then*

$$f(y) = f(x) + \nabla f(tx + (1 - t)y)^T (y - x)$$

for some $t \in [0, 1]$.

Proof. Consider $g(z) = f(zy + (1 - z)x)$. If f is differentiable then so is g. Then by the Mean Value Theorem, for some $t \in (0, 1)$ we have $g(1) - g(0) = g'(t)$. By the chain rule,

$$g'(t) = \nabla f(x + t(y - x))^T (y - x)$$

Using $g(1) = f(y)$ and $g(0) = f(x)$, we have

$$\iff \nabla f(tx + (1 - t)y)^T (y - x) = g(1) - g(0) = f(y) - f(x)$$

$\square$

Taylor's Theorem (second order). Consider a function $f : \mathbb{R}^n \to \mathbb{R}$. Let $a \in \mathbb{R}^n$ be a fixed point. Then Taylor's Theorem states:

If $f(x)$ is twice differentiable on an open ball B around a and $x \in B$ then

$$f(x) = f(a) + (x-a)^T \nabla f(a) + \frac{1}{2}(x-a)^T \nabla^2 f(b)(x-a)$$

for some b on the line segment joining a and x.

This can also be expressed as follows. *Let $x, y \in \mathbb{R}^n$. If $f(x)$ is twice continuously differentiable, then*

$$f(y) = f(x) + \nabla f(x)^T (y-x) + \frac{1}{2}(y-x)^T \nabla^2 f(ty + (1-t)x)(y-x)$$

for some $t \in [0,1]$.

Proof. Consider $g(z) = f(zy + (1-z)x)$. If f is differentiable then so is g. Then by the second order case of Taylor's Theorem in one dimension, for some $t \in (0,1)$ we have $g(1) = g(0) + g'(0) + (1/2)g''(t)$. By the chain rule,

$$g''(t) = \frac{\partial}{\partial t} \nabla f(x + t(y-x))^T (y-x) = (y-x)^T \nabla^2 f(x + t(y-x))^T (y-x)$$

Using this result along with $g(1) = f(y)$, $g(0) = f(x)$, and $g'(0) = \nabla f(x)^T (y-x)$, we have

$$f(y) = f(x) + \nabla f(x)^T (y-x) + \frac{1}{2}(y-x)^T \nabla^2 f(x + t(y-x))^T (y-x)$$

□

2.4 Optimizing functions of several variables

Functions of two variables [edit]

Suppose that $f(x, y)$ is a differentiable real function of two variables whose second partial derivatives exist. The Hessian matrix H of f is the 2 × 2 matrix of partial derivatives of f:

$$H(x,y) = \begin{pmatrix} f_{xx}(x,y) & f_{xy}(x,y) \\ f_{yx}(x,y) & f_{yy}(x,y) \end{pmatrix}.$$

Define $D(x, y)$ to be the determinant

$$D(x,y) = \det(H(x,y)) = f_{xx}(x,y) f_{yy}(x,y) - \left(f_{xy}(x,y)\right)^2,$$

of H. Finally, suppose that (a, b) is a critical point of f (that is, $f_x(a, b) = f_y(a, b) = 0$). Then the second partial derivative test asserts the following:[1]

1. If $D(a,b) > 0$ and $f_{xx}(a,b) > 0$ then (a,b) is a local minimum of f.
2. If $D(a,b) > 0$ and $f_{xx}(a,b) < 0$ then (a,b) is a local maximum of f.
3. If $D(a,b) < 0$ then (a,b) is a saddle point of f.
4. If $D(a,b) = 0$ then the second derivative test is inconclusive, and the point (a, b) could be any of a minimum, maximum or saddle point.

Functions of many variables [edit]

For a function f of two or more variables, there is a generalization of the rule above. In this context, instead of examining the determinant of the Hessian matrix, one must look at the eigenvalues of the Hessian matrix at the critical point. The following test can be applied at any critical point (a, b, ...) for which the Hessian matrix is invertible:

1. If the Hessian is positive definite (equivalently, has all eigenvalues positive) at (a, b, ...), then f attains a local minimum at (a, b, ...).
2. If the Hessian is negative definite (equivalently, has all eigenvalues negative) at (a, b, ...), then f attains a local maximum at (a, b, ...).
3. If the Hessian has both positive and negative eigenvalues then (a, b, ...) is a saddle point for f (and in fact this is true even if (a, b, ...) is degenerate).

2.5 Lagrange Multipliers

: to flesh out! http://tutorial.math.lamar.edu/Classes/CalcIII/LagrangeMultipliers.aspx

2.6 Line Integrals

(p. 555 of Strang book)

Suppose a force in two-dimensional space is given by $\boldsymbol{F} = M\boldsymbol{i} + N\boldsymbol{j}$. Then the work done by this force on a particle moving along a curve C is given by

$$W = \int_C \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{R} = \int_C M\mathrm{d}x + N\mathrm{d}y$$

Along a curve in three-dimensional space the work done by a three-dimensional force $\boldsymbol{F} = M\boldsymbol{i} + N\boldsymbol{j} + P\boldsymbol{j}$ is given by

$$W = \int_C \boldsymbol{F} \cdot \boldsymbol{T}\mathrm{d}s = \int_C \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{R} = \int_C M\mathrm{d}x + N\mathrm{d}y + P\mathrm{d}z$$

where the tangent vector $\boldsymbol{T}$ is given by

$$\boldsymbol{T} = \frac{\mathrm{d}\boldsymbol{R}}{\mathrm{d}s}$$

Green's Theorem: Suppose the region R is bounded by the simple closed piecewise smooth curve C. Then an integral over R equals a line integral around C:

$$\oint_C \boldsymbol{F} \cdot \mathrm{d}\boldsymbol{R} = \oint_C M\mathrm{d}x + N\mathrm{d}y = \int\int_R \left(\frac{\partial N}{\partial x} - \frac{\partial M}{\partial y}\right)\mathrm{d}x\ \mathrm{d}y$$

Line integrals chapter! http://tutorial.math.lamar.edu/Classes/CalcIII/LineIntegralsIntro.aspx

Surface inegrals chapter! http://tutorial.math.lamar.edu/Classes/CalcIII/SurfaceIntegralsIntro.aspx

2.7 Miscellaneous

13A The tangent plane at (x_0, y_0, z_0) has the same slopes as the surface $z = f(x, y)$. The equation of the tangent plane (a linear equation) is

$$z - z_0 = \left(\frac{\partial f}{\partial x}\right)_0 (x - x_0) + \left(\frac{\partial f}{\partial y}\right)_0 (y - y_0). \quad (1)$$

The normal vector $\mathbf{N}$ to that plane has components $(\partial f/\partial x)_0$, $(\partial f/\partial y)_0$, -1.

13B The tangent plane to the surface $F(x, y, z) = c$ has the linear equation

$$\left(\frac{\partial F}{\partial x}\right)_0 (x - x_0) + \left(\frac{\partial F}{\partial y}\right)_0 (y - y_0) + \left(\frac{\partial F}{\partial z}\right)_0 (z - z_0) = 0. \quad (7)$$

The normal vector is $\mathbf{N} = \left(\frac{\partial F}{\partial x}\right)_0 \mathbf{i} + \left(\frac{\partial F}{\partial y}\right)_0 \mathbf{j} + \left(\frac{\partial F}{\partial z}\right)_0 \mathbf{k}$.

$$dz = (\partial z/\partial x)_0 dx + (\partial z/\partial y)_0 dy \quad \text{or} \quad df = f_x dx + f_y dy. \quad (10)$$

This is the ***total differential***. All letters dz and df and dw can be used, but ∂z and ∂f are not used. Differentials suggest small movements in x and y; then dz is the resulting movement in z. On the tangent plane, equation (10) holds exactly.

The **directional derivative**, denoted $D_v f(x, y)$, is a derivative of a multivariable function in the direction of a vector $\boldsymbol{v}$. It is the scalar projection of the gradient onto $\boldsymbol{v}$.

$$D_v f(x, y) = \text{comp}_v \nabla f(x, y) = \frac{\nabla f(x, y) \cdot \boldsymbol{v}}{|\boldsymbol{v}|}$$

2.8 Practice Problems

13F The directional derivative is $D_{\mathbf{u}} f = (\text{grad}\, f) \cdot \mathbf{u}$. The level direction is perpendicular to grad f, since $D_{\mathbf{u}} f = 0$. ***The slope $D_{\mathbf{u}} f$ is largest when u is parallel to*** grad f. That maximum slope is the length $|\text{grad}\, f| = \sqrt{f_x^2 + f_y^2}$:

$$\textit{for} \quad \mathbf{u} = \frac{\text{grad}\, f}{|\text{grad}\, f|} \quad \textit{the slope is} \quad (\text{grad}\, f) \cdot \mathbf{u} = \frac{|\text{grad}\, f|^2}{|\text{grad}\, f|} = |\text{grad}\, f|.$$

$$\int_C g(x, y)\, ds = \text{limit of} \sum_{i=1}^{N} g(x_i, y_i) \Delta s_i \quad \text{as} \quad (\Delta s)_{\max} \to 0.$$

The differential ds becomes $(ds/dt)dt$. Everything changes over to t:

$$\int g(x, y)ds = \int_{t=a}^{t=b} g(x(t), y(t))\sqrt{(dx/dt)^2 + (dy/dt)^2}\, dt.$$

19. Let f and g be twice-differentiable real-valued functions defined on $\mathbb{R}$. If $f'(x) > g'(x)$ for all $x > 0$, which of the following inequalities must be true for all $x > 0$?

(A) $f(x) > g(x)$

(B) $f''(x) > g''(x)$

(C) $f(x) - f(0) > g(x) - g(0)$

(D) $f'(x) - f'(0) > g'(x) - g'(0)$

(E) $f''(x) - f''(0) > g''(x) - g''(0)$

Solution 19. (C) There is no reason that $f(x) > g(x)$, or that $f''(x) > g''(x)$. But we do know that

$$\int_0^x f'(t)\,dt > \int_0^x g'(t)\,dt \implies f(x) - f(0) > g(x) - g(0).$$

This is precisely an answer.

22. What is the volume of the solid in xyz-space bounded by the surfaces $y = x^2$, $y = 2 - x^2$, $z = 0$, and $z = y + 3$?

(A) $\frac{8}{3}$ (B) $\frac{16}{3}$ (C) $\frac{32}{3}$ (D) $\frac{104}{105}$ (E) $\frac{208}{105}$

Solution 22. (C) It looks like our x-coordinates are running over $[-1, 1]$, with y depending on x and z depending on y. To find the volume of the solid, we just need to integrate the constant function 1. We must therefore compute

$$\begin{aligned}\int_{-1}^{1}\int_{x^2}^{2-x^2}\int_0^{y+3} 1\,dz\,dy\,dx &= \int_{-1}^{1}\int_{x^2}^{2-x^2} y + 3\,dy\,dx\\ &= \int_{-1}^{1} \left((2-x^2)^2/2 + 3(2-x^2)\right) - \left((x^2)^2/2 + 3(x^2)\right)\,dx\\ &= \int_{-1}^{1} 8 - 8x^2\,dx\\ &= 8x - 8x^3/3\Big|_{-1}^{1} = (8 - 8/3) - (-8 + 8/3) = 32/3.\end{aligned}$$

24. Let h be the function defined by $h(x) = \int_0^{x^2} e^{x+t}\,dt$ for all real numbers x. Then $h'(1) =$

(A) $e - 1$ (B) e^2 (C) $e^2 - e$ (D) $2e^2$ (E) $3e^2 - e$

Solution 24. (E) We can actually just integrate this, and not worry about differentiation under the integral.

$$\int_0^{x^2} e^{x+t}\,dt = e^x \int_0^{x^2} e^t\,dt = e^x(e^{x^2}-1) = e^{x^2+x} - e^x.$$

Then deriving that,

$$h'(x) = (2x+1)e^{x^2+x} - e^x,$$

whence our result follows immediately.

26. Let $f(x, y) = x^2 - 2xy + y^3$ for all real x and y. Which of the following is true?

(A) f has all of its relative extrema on the line $x = y$.

(B) f has all of its relative extrema on the parabola $x = y^2$.

(C) f has a relative minimum at $(0, 0)$.

(D) f has an absolute minimum at $\left(\frac{2}{3}, \frac{2}{3}\right)$.

(E) f has an absolute minimum at $(1, 1)$.

Solution 26. (A) We are concerned about its extrema, we should find some partial derivatives.

$$f_x = 2x - 2y, \quad f_y = -2x + 3y^2.$$

We would like to know when they are both zero. The first equation gives us $x = y$ and the second gives us $2x = 3y^2$, so that

$$2y = 3y^2 \implies (3y-2)y = 0 \implies y = 0, 2/3.$$

Therefore our solutions are $(0, 0)$ and $(2/3, 2/3)$. Indeed, our relative extrema are all on the line $x = y$. To do some more checking (which you should not do on the actual test),

$$f_{xx} = 2, \quad f_{yy} = 6y, \quad f_{xy} = f_{yx} = -2.$$

Then the determinant of the Hessian is $12y - 4$. This shows that $(0, 0)$ is a saddle point. There is no reason that $(2/3, 2/3)$ is an absolute minimum without further verification, and $(1, 1)$ needn't be an extreme point.

27. Consider the two planes $x + 3y - 2z = 7$ and $2x + y - 3z = 0$ in $\mathbb{R}^3$. Which of the following sets is the intersection of these planes?

(A) $\varnothing$

(B) $\{(0, 3, 1)\}$

(C) $\{(x, y, z) : x = t,\ y = 3t,\ z = 7 - 2t,\ t \in \mathbb{R}\}$

(D) $\{(x, y, z) : x = 7t,\ y = 3 + t,\ z = 1 + 5t,\ t \in \mathbb{R}\}$

(E) $\{(x, y, z) : x - 2y - z = -7\}$

Solution 27. (D) First, we know that the intersection of two planes in $\mathbb{R}^3$ should be either a plane or a line. In our case, the two planes are definitely not the same, so we will obtain a line. The slope of the line can be found by taking the cross product of the normal vectors of the two planes in question.

$$(1,3,-2)\times(2,1,-3) = \det\begin{bmatrix} i & j & k \\ 1 & 3 & -2 \\ 2 & 1 & -3 \end{bmatrix} = (-7,-1,-5).$$

The only solution corresponding to this slope is (D), as the coefficients of t in (x,y,z) are $(7,1,5)$.

32. $\dfrac{d}{dx}\displaystyle\int_{x^3}^{x^4} e^{t^2}\,dt =$

(A) $e^{x^6}\left(e^{x^8-x^6}-1\right)$ (B) $4x^3e^{x^8}$ (C) $\dfrac{1}{\sqrt{1-e^{x^2}}}$ (D) $\dfrac{e^{x^2}}{x^2}-1$ (E) $x^2e^{x^6}\left(4xe^{x^8-x^6}-3\right)$

Solution 32. (E) We can sort this out in two steps and apply the fundamental theorem to each.

$$\frac{d}{dx}\left(\int_{x^3}^{0} e^{t^2}\,dx + \int_0^{x^4} e^{t^2}\,dx\right)$$

For the first,

$$\frac{d}{dx}\int_{x^3}^{0} e^{t^2}\,dx = -\frac{d}{dx}\int_0^{x^3} e^{t^2}\,dx = -3x^2e^{x^6}$$

For the second,

$$\frac{d}{dx}\int_0^{x^4} e^{t^2}\,dx = 4x^3e^{x^8}.$$

All told, our integral is $x^2e^{x^6}(4xe^{x^8-x^6}-3)$.

41. Let ℓ be the line that is the intersection of the planes $x+y+z=3$ and $x-y+z=5$ in $\mathbb{R}^3$. An equation of the plane that contains $(0,0,0)$ and is perpendicular to ℓ is

(A) $x-z=0$

(B) $x+y+z=0$

(C) $x-y-z=0$

(D) $x+z=0$

(E) $x+y-z=0$

Solution 41. (A) The first plane is determined by the normal vector $(1, 1, 1)$, and the second determined by $(1, -1, 1)$. Therefore the slope of ℓ is determined by a vector perpendicular to those, i.e. the cross product.

$$(1,1,1) \times (1,-1,1) = \det \begin{bmatrix} i & j & k \\ 1 & 1 & 1 \\ 1 & -1 & 1 \end{bmatrix} = (2, 0, -2).$$

41. Let C be the circle $x^2 + y^2 = 1$ oriented counterclockwise in the xy-plane. What is the value of the line integral $\oint_C (2x - y)\,dx + (x + 3y)\,dy$?

(A) 0 (B) 1 (C) $\frac{\pi}{2}$ (D) π (E) 2π

Solution 41. (E) This is a classic Green's theorem problem.

$$\oint_{\partial D} L\,dx + M\,dy = \iint_D \left(\frac{\partial M}{\partial x} - \frac{\partial L}{\partial y} \right) dx\,dy.$$

In our case,

$$\oint_C (2x - y)\,dx + (x + 3y)\,dy = \iint_D (1 + 1)\,dx\,dy = 2A,$$

where A is the area of the unit circle, i.e. π.

So that is the slope of ℓ. We need this to be the normal vector for the plane in question, so it seems that $(1, 0, -1)$ is our best bet (out of the given options).

$$y' + xy = x$$
$$y(0) = -1$$

44. If y is a real-valued function defined on the real line and satisfying the initial value problem above, then $\lim_{x \to -\infty} y(x) =$

(A) 0 (B) 1 (C) -1 (D) ∞ (E) $-\infty$

Solution 44. (B) Putting it in simpler terms,

$$\frac{dy}{dx} + xy = x \implies \frac{dy}{dx} = x(1 - y) \implies \frac{dy}{1 - y} = x\,dx.$$

Integrating both sides, we obtain

$$-\log(1 - y) = x^2/2 + C' \implies 1 - y = Ce^{-x^2/2} \implies y = 1 - Ce^{-x^2/2}.$$

Solving the initial value problem gives $C = 2$. Furthermore, as $x \to -\infty$, the second term above vanishes so we get 1 in the limit.

48. Let g be the function defined by $g(x, y, z) = 3x^2y + z$ for all real x, y, and z. Which of the following is the best approximation of the directional derivative of g at the point $(0, 0, \pi)$ in the direction of the vector $\mathbf{i} + 2\mathbf{j} + 3\mathbf{k}$? (Note: $\mathbf{i}$, $\mathbf{j}$, and $\mathbf{k}$ are the standard basis vectors in $\mathbb{R}^3$.)

(A) 0.2 (B) 0.8 (C) 1.4 (D) 2.0 (E) 2.6

Solution 48. (B) It would be good to recall the formula for the directional derivative. We take the gradient of the function then take its scalar product with the normalised vector in the direction we want. To begin,

$$\nabla g = (6xy, 3x^2, 1).$$

At the point $(0, 0, \pi)$, we have $\nabla g = (0, 0, 1)$. That works out pretty well for us. The normalised version of the vector $(1, 2, 3)$ is $(1/\sqrt{14}, 2/\sqrt{14}, 3/\sqrt{14})$. Dotting this with $(0, 0, 1)$ gives $3/\sqrt{14}$, and since $\sqrt{14} = 3.5$ or so our answer should be closer to 0.8 than 0.2.

48. Consider the theorem: If f and f' are both strictly increasing real-valued functions on the interval $(0, \infty)$, then $\lim_{x\to\infty} f(x) = \infty$. The following argument is suggested as a proof of this theorem.

(1) By the Mean Value Theorem, there is a c_1 in the interval $(1, 2)$ such that

$$f'(c_1) = \frac{f(2) - f(1)}{2 - 1} = f(2) - f(1) > 0.$$

(2) For each $x > 2$, there is a c_x in $(2, x)$ such that $\frac{f(x) - f(2)}{x - 2} = f'(c_x)$.

(3) For each $x > 2$, $\frac{f(x) - f(2)}{x - 2} = f'(c_x) > f'(c_1)$ since f' is strictly increasing.

(4) For each $x > 2$, $f(x) > f(2) + (x - 2)f'(c_1)$.

(5) $\lim_{x\to\infty} f(x) = \infty$

Which of the following statements is true?

(A) The argument is valid.

(B) The argument is not valid since the hypotheses of the Mean Value Theorem are not satisfied in (1) and (2).

(C) The argument is not valid since (3) is not valid.

(D) The argument is not valid since (4) cannot be deduced from the previous steps.

(E) The argument is not valid since (4) does not imply (5).

Solution 48. (A) The only issue here seems to be that (4) implies that $f(x)$ gets very large so long as $f'(c_1)$ is positive. But we know that it is, since f is a strictly increasing function. Therefore everything is satisfactory.

3 Differential Equations

61. A tank initially contains a salt solution of 3 grams of salt dissolved in 100 liters of water. A salt solution containing 0.02 grams of salt per liter of water is sprayed into the tank at a rate of 4 liters per minute. The sprayed solution is continually mixed with the salt solution in the tank, and the mixture flows out of the tank at a rate of 4 liters per minute. If the mixing is instantaneous, how many grams of salt are in the tank after 100 minutes have elapsed?

    (A) 2 (B) $2 - e^{-2}$ (C) $2 + e^{-2}$ (D) $2 - e^{-4}$ (E) $2 + e^{-4}$

Solution 61. (E) We can set this up as a differential equation. Let s denote the amount of salt in the tank, and let t denote time. We have the initial condition of $s(0) = 3$. $s'(t)$ depends on two factors: the salt flowing in and the salt flowing out. The salt flows in constantly at a rate of 0.08 grams per minute, and the salt flows out at a rate of $4 \cdot (s/100) = s/25$ grams per minute. Therefore

$$s'(t) = \frac{ds}{dt} = 0.08 - s(t)/25 \implies \frac{ds}{dt} = 0.04(2-s) \implies \frac{ds}{2-s} = 0.04\,dt.$$

Doing the usual calculus,

$$-\log(2-s) = 0.04t + C' \implies 2 - s = Ce^{-0.04t} \implies s(t) = 2 - Ce^{-0.04t}.$$

The initial condition tells us that $C = -1$, so $s(t) = 2 + e^{-0.04t}$. Plugging in $t = 100$ gives our answer.

4 Real Analysis

These are my notes from Math 4650: Analysis I at Cal State LA.

4.1 Midterm 1

4.1.1 Homework 1

Definition: Let $S \subseteq \mathbb{R}$. We say that S is **bounded from above** if $\exists\, b \in \mathbb{R}$ where

$$s \leq b \;\forall\; s \in S$$

If this is the case, we call b an **upper bound** of S.

If $b \leq c$ for all upper bounds c of S, we call b the **supremum** of S: $b = \sup(S)$.

We say that S is **bounded from below** if $\exists\, a \in \mathbb{R}$ where

$$s \geq a \;\forall\; s \in S$$

If this is the case, we call a a **lower bound** of S.

If $a \geq d$ for all lower bounds d of S, we call a the **infimum** of S: $a = \inf(S)$.

Useful Sup/Inf Fact: Let $S \in \mathbb{R}$, $S \neq \emptyset$.

(1) Suppose S is bounded from above by an element b. Then $b = \sup(S) \iff \forall\, \epsilon > 0\; \exists\, x \in S$ with

$$b - \epsilon < x \leq b$$

(2) Suppose S is bounded from below by an element a. Then $a = \inf(S) \iff \forall\, \epsilon > 0\; \exists\, x \in S$ with

$$a \leq x < a + \epsilon$$

Completeness Axiom: Let S be a nonempty subset of $\mathbb{R}$. If S is bounded from above, then $\sup(S)$ exists. If S is bounded from below, then $\inf(S)$ exists.

Facts about absolute value:

(1) $|x - y| < \epsilon \iff y - \epsilon < x < y + \epsilon$ (proof: in notes 08/23)

(2) $|ab| = |a||b|$ (proof: 7(c) in Homework 1)

(3) Let $\epsilon > 0$. Then $|a| < \epsilon \iff -\epsilon < a < \epsilon$. (Proof: follows from (1) if $x = a$, $y = 0$.)

(4) $-|a| \leq a \leq |a|$ (proof: Follows from (1) if $x = a$, $y = 0$, $\epsilon = |a|$.)

(5) **Triangle Inequality:** $|a + b| \leq |a| + |b|$ (Proof in notes 08/23)

(6) $|\;|a| - |b|\;| \leq |a - b|$ (Proof: 7(d) in Homework 1)

(7) **Triangle Inequality:** $|a - b| \leq |a| + |b|$ (Proof: follows from (5), let $b = -b$.)

(8) If $a < x < b$ and $a < y < b$ then $|x - y| < b - a$. (Proof: 7(a) in Homework 1)

(9) $|a - b| = |b - a|$ (Proof: 7(b) in Homework 1.)

4.1.2 Homework 2

Definition: A sequence (a_n) of real numbers is said to **converge** to a **limit** $L \in \mathbb{R}$ if $\forall\ \epsilon > 0\ \exists\ N > 0$ where

$$n \geq N \implies |a_n - L| < \epsilon$$

We say that (a_n) **diverges** if it does not converge.

Definition: A sequence (a_n) of real numbers is **bounded** if $\exists\ M > 0$ where $\forall\ n \in \mathbb{N}$

$$|a_n| \leq M$$

Theorem. If (a_n) converges then (a_n) is bounded.

Definition: Let (a_n) be a sequence of real numbers. We say that (a_n) is a **Cauchy sequence** if $\forall\ \epsilon > 0\ \exists\ N$ where

$$n, m \geq N \implies |a_n - a_m| < \epsilon$$

Theorem. (a_n) is Cauchy if and only if (a_n) converges.

Corollary. If (a_n) is Cauchy then (a_n) is bounded.

4.2 Midterm 2

4.2.1 Homework 3

Limits of functions at infinity. Let f be a real-valued function defined on some set D where D contains an interval of the form (a, ∞). Let $L \in \mathbb{R}$. We say

$$\lim_{x \to \infty} f(x) = L$$

if $\forall\ \epsilon > 0\ \exists\ N \in \mathbb{R}$ where

$$x \geq N \implies |f(x) - L| < \epsilon$$

Definition: Let $D \subseteq \mathbb{R}$. Let $a \in \mathbb{R}$. We say that a is a **limit point** (or "cluster point," or "accumulation point") of D if $\forall\ \delta > 0\ \exists\ x \in D$ where

$$x \neq a \text{ and } |x - a| < \delta$$

(Note that a may or may not be contained in D.)

Limit of a function at a: Let $D \subseteq \mathbb{R}$ and $f : d \to \mathbb{R}$. Let a be a limit point of D. Let $x \in D$. We say that f has a *limit as x tends to a* if $\exists\ L \in \mathbb{R}$ where $\forall\ \epsilon > 0\ \exists\ \delta > 0$ such that

$$0 < |x - a| < \delta \implies |f(x) - L| < \epsilon$$

and we write

$$\lim_{x \to a} f(x) = L$$

Properties of Limits: Let $D \in \mathbb{R}$ and let a be a limit point of D. Suppose $f : D \to \mathbb{R}$ and $g : D \to \mathbb{R}$. Let $\alpha \in \mathbb{R}$.

(1) If $\lim_{x \to a} f(x) = L$ and $\lim_{x \to a} g(x) = M$ then

(a)
$$\lim_{x \to a} \alpha = \alpha$$

(b)
$$\lim_{x \to a} [f(x) + g(x)] = L + M$$

(c)
$$\lim_{x \to a} [f(x) - g(x)] = L - M$$

(d)
$$\lim_{x \to a} [f(x) \cdot g(x)] = L \cdot M$$

(e)
$$\lim_{x \to a} [\alpha \cdot f(x)] = \alpha \cdot L$$

(2) If $h : D \to \mathbb{R}$ and $h(x) \neq 0 \ \forall \ x \in D$ and $\lim_{x\to a} h(x) = H \neq 0$, then

$$\lim_{x\to a} \frac{1}{h(x)} = \frac{1}{H}$$

Note that properties (2) and (1)(d) combined imply

$$\lim_{x\to a} \frac{f(x)}{h(x)} = \frac{L}{H}$$

4.2.2 Homework 4

Continuity: Let $D \subseteq \mathbb{R}$ and $f : D \to \mathbb{R}$ and $a \in D$. Then f is **continuous** at a if $\lim_{x\to a} f(x)$ exists and

$$\lim_{x\to a} f(x) = f(a)$$

(Note: if f is continuous at a, then we can say $\forall \ \epsilon > 0 \ \exists \ \delta > 0$ such that

$$|x - a| < \delta \implies |f(x) - L| < \epsilon$$

that is, we don't need to say $0 < |x - a| < \delta$).

If $B \subseteq D$, then f is **continuous on B** if f is continuous at every $b \in B$.

Intermediate Value Theorem: Let f be continuous on $[a, b]$ and suppose $f(a) < f(b)$. $\forall \ d$ such that

$$f(a) < d < f(b)$$

$\exists \ c \in \mathbb{R}$ where

$$a < c < b, \ \ f(c) = d$$

4.3 Final

4.3.1 Homework 5

Definition: Let $S \subseteq \mathbb{R}$. We say $x \in \mathbb{R}$ is an **interior point** of S if there exists an open interval (a, b) where

$$x \in (a, b) \text{ and } (a, b) \subseteq S$$

Open sets: Let $S \subseteq \mathbb{R}$. We say S is **open** if every $x \in S$ is an interior point of S.

Closed sets: Let $S \subseteq \mathbb{R}$. We say S is **closed** if $\mathbb{R} \setminus S$ is open.

Theorem. A set is closed if and only if it contains all of its limit points.

Facts about open and closed sets: Suppose $a, b \in \mathbb{R}$. Then

- (a, ∞) is open (Proof: Homework 5 problem 5b).
- $(-\infty, b)$ is open (Proof: Homework 5 problem 5a).
- (a, b) is open (Proof: class notes).
- If $a < b$, then $[a, b]$ is closed (Proof: Homework 5 problem 5c).
- If A and B are open, then $A \cup B$ and $A \cap B$ are open (Proof: Homework 5 problem 3).
- If A and B are closed, then $A \cup B$ and $A \cap B$ are closed (Proof: Homework 5 problem 4).
- $\mathbb{R}$ is open (Proof: Homework 5 problem 1) and closed (Proof: $\mathbb{R} \setminus \mathbb{R} = \emptyset$ is open).
- $\emptyset$ is open (Proof: Homework 5 problem 2) and closed (Proof: $\mathbb{R} \setminus \emptyset = \mathbb{R}$ is open).

Definition: Let $S \subseteq \mathbb{R}$. An **open cover** of S is a collection $X = \{\mathcal{O}_\alpha \mid \alpha \in I\}$ where each set $\mathcal{O}_\alpha$ is an open subset of $\mathbb{R}$ such that

$$S \subseteq \bigcup_{\alpha \in I} \mathcal{O}_\alpha$$

(Here I is some set that indexes the $\mathcal{O}_\alpha$).

If $X' \subseteq X$ such that

$$S \subseteq \bigcup_{\mathcal{O}_\alpha \in X'} \mathcal{O}_\alpha$$

then X' is called a **subcover** of S contained in X. In addition, if X' is finite then we call X' a **finite subcover** of S contained in X.

Compactness: Let $S \subseteq \mathbb{R}$. We say that S is **compact** if every open cover of S contains a finite subcover.

Definition: Let $S \subseteq \mathbb{R}$. We say that S is **bounded** if $\exists\ M > 0$ where $S \subseteq [-M, M]$.

Note: S is bounded if and only if $|s| \leq M\ \forall\ s \in S$.

Heine-Borel Theorem. Let $S \subseteq \mathbb{R}$. S is compact if and only if S is closed and bounded.

Theorem. Let $f : D \to \mathbb{R}$ be continuous on D. If $X \subseteq D$ and X is compact (closed and bounded), then

$$f(\bar{x}) = \{f(x) \mid x \in X\}$$

is compact (closed and bounded).

Corollary: Suppose $f : D \to \mathbb{R}$ where D is closed and bounded. Then there exists $a, b \in D$ where $f(a)$ is the min of f on D and $f(b)$ is the max of f on D.

4.3.2 Homework 6

Uniform Continuity: Let $D \subseteq \mathbb{R}$ and let $f : D \to \mathbb{R}$. We say that f is **uniformly continuous** on D if $\forall\ \epsilon > 0\ \exists\ \delta > 0$ where

$$x, y \in D \text{ and } 0 < |x - y| < \delta \implies |f(x) - f(y)| < \epsilon$$

Uniform continuity implies continuity. Suppose $f : D \to \mathbb{R}$ where $D \subseteq \mathbb{R}$. If f is uniformly continuous on D, then f is continuous at every $a \in D$.

4.4 Problems from Practice Math GRE Subject Tests

38. Let A and B be nonempty subsets of $\mathbb{R}$ and let $f : A \to B$ be a function. If $C \subseteq A$ and $D \subseteq B$, which of the following must be true?

(A) $C \subseteq f^{-1}\big(f(C)\big)$

(B) $D \subseteq f\big(f^{-1}(D)\big)$

(C) $f^{-1}\big(f(C)\big) \subseteq C$

Solution 38. (A) Neither of the equalities should hold – these are in fact nonsense statements, as one side lies in A and the other in B. To unravel the remaining two sets,

$$f^{-1}(f(C)) = \{x \in A : f(x) \in f(C)\}, \quad f(f^{-1}(D)) = f(\{y \in A : f(y) \in D\})$$

Clearly the second set must always be contained in D, but not the other way around. Similarly the first set certainly contains all $c \in C$ (as $f(c) \in f(C)$) but not the other way around.

47. The function $f: \mathbb{R} \to \mathbb{R}$ is defined as follows.

$$f(x) = \begin{cases} 3x^2 & \text{if } x \in \mathbb{Q} \\ -5x^2 & \text{if } x \notin \mathbb{Q} \end{cases}$$

Which of the following is true?

(A) f is discontinuous at all $x \in \mathbb{R}$.

(B) f is continuous only at $x = 0$ and differentiable only at $x = 0$.

(C) f is continuous only at $x = 0$ and nondifferentiable at all $x \in \mathbb{R}$.

(D) f is continuous at all $x \in \mathbb{Q}$ and nondifferentiable at all $x \in \mathbb{R}$.

(E) f is continuous at all $x \notin \mathbb{Q}$ and nondifferentiable at all $x \in \mathbb{R}$.

Solution 47. (B) A classic kind of problem. We are clearly continuous and differentiable at 0. Anywhere else, near a rational number there is an irrational number and vice versa. Therefore there can be no continuity anywhere but at 0, and hence no differentiability either.

57. For each positive integer n, let x_n be a real number in the open interval $\left(0, \frac{1}{n}\right)$. Which of the following statements must be true?

I. $\lim_{n\to\infty} x_n = 0$

II. If f is a continuous real-valued function defined on $(0, 1)$, then $\{f(x_n)\}_{n=1}^{\infty}$ is a Cauchy sequence.

III. If g is a uniformly continuous real-valued function defined on $(0, 1)$, then $\lim_{n\to\infty} g(x_n)$ exists.

(A) I only (B) I and II only (C) I and III only (D) II and III only (E) I, II, and III

Solution 57. (C) I is true, since $\lim_{n\to\infty} x_n$ must be bounded between 0 and $\lim_{n\to\infty} 1/n = 0$. Unfortunately, x_n does not converge inside $(0, 1)$. There is no reason therefore that $f(x_n)$ should be a convergent sequence – suppose that $f(x) = 1/x$, so that $f(x_n)$ is certainly not Cauchy. However, if g is uniformly continuous, then g extends to a continuous function on $[0, 1]$. Now x_n is a convergent sequence, so $\lim_{n\to\infty} g(x_n) = g(\lim_{n\to\infty} x_n) = g(0)$ exists.

60. A real-valued function f defined on $\mathbb{R}$ has the following property.

For every positive number ϵ, there exists a positive number δ such that
$$|f(x) - f(1)| \geq \epsilon \text{ whenever } |x-1| \geq \delta.$$

This property is equivalent to which of the following statements about f?

(A) f is continuous at $x = 1$.

(B) f is discontinuous at $x = 1$.

(C) f is unbounded.

(D) $\lim_{|x|\to\infty} |f(x)| = \infty$

(E) $\int_0^\infty |f(x)|dx = \infty$

Solution 60. (D) While it looks like this is the opposite of continuity, that should read 'there exists $\varepsilon > 0$'. What the statement says is that we not only get arbitrarily far away from $f(1)$, but we must for all x sufficiently far away from 1. So as $|x|$ gets very large, so does $|f(x)|$.

63. For any nonempty sets A and B of real numbers, let $A \cdot B$ be the set defined by
$$A \cdot B = \{xy : x \in A \text{ and } y \in B\}.$$

If A and B are nonempty bounded sets of real numbers and if $\sup(A) > \sup(B)$, then $\sup(A \cdot B) =$

(A) $\sup(A)\sup(B)$

(B) $\sup(A)\inf(B)$

(C) $\max\{\sup(A)\sup(B),\ \inf(A)\inf(B)\}$

(D) $\max\{\sup(A)\sup(B),\ \sup(A)\inf(B)\}$

(E) $\max\{\sup(A)\sup(B),\ \inf(A)\sup(B),\ \inf(A)\inf(B)\}$

Solution 63. (E) The supremum is either going to be the product of the two largest positive numbers in A and B or the product of the two smallest negative numbers in A and B. That means we should look for $\sup \cdot \sup$ or $\inf \cdot \inf$. However, it might be the case that B contains only negative numbers and A contains only positive numbers. Then the largest value in $A \cdot B$ will be attained by the smallest positive element of A and the largest negative element of B, giving us our third option: $\inf A \cdot \sup B$.

5 Probability

These are my notes from taking Math 505A at USC and the textbook *Probability and Random Processes* (Grimmet and Stirkazer) 3rd edition.

5.1 To Know for Math 505A Midterm 1 (Discrete Random Variables)

5.1.1 Definitions

Definition 5.1. The **probability mass function** of a discrete random variable X is the function $f : \mathbb{R} \to [0,1]$ given by $f(x) = \Pr(X = x)$.

Definition 5.2. The **(cumulative) distribution function** of a discrete random variable F is given by

$$F(x) = \sum_{i: x_i \leq x} f(x_i)$$

Definition 5.3. The **joint probability mass function** $f : \mathbb{R}^2 \to [0,1]$ of two discrete random variables X and Y is given by

$$f(x, y) = \Pr(X = x \cap Y = y)$$

Definition 5.4. The **joint distribution function** $F : \mathbb{R}^2 \to [0,1]$ is given by

$$F(x, y) = \Pr(X \leq x \cap Y \leq y)$$

Definition 5.5. If $\Pr(B) > 0$ then the **conditional probability** that A occurs given that B occurs is defined to be

$$\Pr(A \mid B) = \frac{\Pr(A \cap B)}{\Pr(B)}$$

Definition 5.6. Two random variables X and Y are **independent** if and only if $\Pr(X \cap Y) = \Pr(X)\Pr(Y)$.

Theorem 1. (Law of total probability). If X is a random variable and Y is a discrete random variable taking on values $y_1, y_2, \ldots, y_n$, then $\Pr(X) = \sum_i \Pr(X \mid Y = y_i) \cdot \Pr(Y = y_i)$. (Can be used to prove independence.)

Definition 5.7. Two random variables X and Y are **uncorrelated** if $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$.

Proposition 2. (a) Two random variables are uncorrelated if and only if their covariance $\text{Cov}(X, Y) = \mathbb{E}\big[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))\big] = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$ equals 0.

(b) If X and Y are independent then they are uncorrelated.

Theorem 3. If X and Y are independent and $g, h : \mathbb{R} \to \mathbb{R}$, then $g(X)$ and $h(Y)$ are also independent.

5.1.2 Conditioning

Definition 5.8. The **conditional distribution function** of Y given $X = x$, written $F_{Y|X}(\cdot \mid x)$, is defined by

$$F_{Y|X}(y \mid x) = \Pr(Y \leq y \mid X = x)$$

Definition 5.9. The **conditional probability mass function** of Y given $X = x$, written $f_{Y|X}(\cdot \mid x)$, is defined by

$$f_{Y|X}(y \mid x) = \Pr(Y = y \mid X = x)$$

Proposition 4. Iterated expectations:

- $\mathbb{E}\big[\mathbb{E}(Y \mid X)\big] = \mathbb{E}(Y)$
- $\mathbb{E}\big[(X \mid Y) \mid Z\big] = \mathbb{E}(X \mid Y)$
- $\mathbb{E}(E(XY \mid Y)) = \mathbb{E}(Y\mathbb{E}(X \mid Y))$

Definition 5.10. Conditional Variance: $\mathrm{Var}(X \mid Y) = \mathbb{E}\big[(X - \mathbb{E}(X \mid Y))^2 \mid Y\big]$

5.1.3 Odds and Ends

Proposition 5. Inclusion-Exclusion Principle:

(a)

$$\Pr\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{k=1}^{n}(-1)^{k+1}\left(\sum_{1 \leq i_1 < \ldots < i_k \leq m} \Pr\left(A_{i1} \cap \ldots \cap A_{ik}\right)\right)$$

(b)

$$\left|\bigcup_{i=1}^{n} A_i\right| = \sum_{k=1}^{n}(-1)^{k+1}\left(\sum_{1 \leq i_1 < \ldots < i_k \leq m} |A_{i1} \cap \ldots \cap A_{ik}|\right)$$

Theorem 6. Sums of random variables. If X and Y are independent then

$$\Pr(X + Y = z) = f_{X+Y}(z) = \sum_x f_X(x) f_Y(z - x) = \sum_y f_X(z - y) f_Y(y)$$

Proposition 7. Variance-Covariance Expansion. Let $X_1, \ldots, X_n$ be random variables. If $\mathbb{E}\left|X_k\right|^2 < \infty$, then

$$\mathrm{Var}(X_1 + \ldots + X_n) = \sum_k \mathrm{Var}(X_k) + \sum_{k \neq m} \sum_m \mathrm{Cov}(X_k, X_m)$$

Proposition 8. (Proposition 1.6.1 in Sheldon Ross *A First Course in Probability.*) There are $\binom{n-1}{r-1}$ distinct positive integer-valued vectors $(x_1, x_2, \ldots, x_r), x_i > 0 \;\forall\, i$ satisfying the equation $x_1 + x_2 + \ldots + x_r = n$.

Proof. (Not rigorous, but a justification.) Imagine we have n indistinguishable objects to allocate to r people. We lay out the n objects and take $r-1$ sticks to place in the $n-1$ spaces between them. The first person gets all the objects to the left of the leftmost stick, the second person gets the objects between the leftmost and second leftmost stick, and so on, until the last person gets all the objects to the right of the rightmost stick. The constraint that x_i be positive is equivalent to saying that each person must receive at least one object. Therefore we must place each stick in a different place. There are $\binom{n-1}{r-1}$ ways to do this.

□

Proposition 9. **(Proposition 1.6.2 in Sheldon Ross *A First Course in Probability.*)**There are $\binom{n+r-1}{r-1}$ distinct nonnegative integer-valued vectors $(x_1, x_2, \dots, x_r), x_i \geq 0 \ \forall \ i$ satisfying the equation $x_1 + x_2 + \ldots + x_r = n$.

Proof. We would like to solve the problem

$$x_1 + x_2 + \ldots + x_r = n, x_i \geq 0 \ \forall i$$

Note that we can transform this problem in the following way:

$$x_1 + 1 + x_2 + 1 + \ldots + x_r + 1 = n + 1 \cdot r, x_i + 1 \geq 1 \ \forall i$$

Letting $y_i = x_i + 1$, we have the equivalent system

$$y_1 + y_2 + \ldots + y_r = n + r, y_i \geq 1 \ \forall i$$

By Proposition 8, the number of distinct solutions to this equation is $\binom{n+r-1}{r-1}$.

□

5.1.4 Methods for Calculating Quantities

- Expectation
  -
    Definition 5.11. $\mathbb{E}(X) = \sum_x x \Pr(X = x)$
  -
    Theorem 10. (a) $\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y)$
    (b) If $X \geq 0$ then $\mathbb{E}(X) \geq 0$
  -
    Theorem 11. **Law of the Unconscious Statistician:** If X has mass function f, and $g : \mathbb{R} \to \mathbb{R}$, then

    $$\mathbb{E}(g(X)) = \sum_x g(x)f(x)$$
  -

Proposition 12. Expectation is a linear operator: $\mathbb{E}\big(\sum_i X_i\big) = \sum_i \mathbb{E}(X_i)$

- Variance
  - **Definition 5.12.** $\mathrm{Var}(X) = \mathbb{E}(X - \mathbb{E}(X))^2$
  - **Proposition 13. (Useful reformulation:)** $\mathrm{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$
  - **Theorem 14. (Some useful results):**
    (a) $\mathrm{Var}(aX) = a^2\mathrm{Var}(X)$
    (b) $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X,Y)$
    (c) $\mathrm{Var}(aX \pm bY) = a^2\mathrm{Var}(X) + b^2\mathrm{Var}(Y) \pm 2ab\mathrm{Cov}(X,Y)$
  - **Theorem 15. Law of Total variance:** $\mathrm{Var}(X) = \mathrm{Var}\big(\mathbb{E}(X \mid Y)\big) + \mathbb{E}\big(\mathrm{Var}(X \mid Y)\big)$
- Covariance
  - **Definition 5.13.** $\mathrm{Cov}(X,Y) = \mathbb{E}\big[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))\big]$
  - **Proposition 16. (Useful reformulation):** $\mathrm{Cov}(X) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$
  - **Definition 5.14.** Conditional covariance:

$$\mathrm{Cov}(X,Y \mid Z) = \mathbb{E}(XY \mid Z) - \mathbb{E}(X \mid Z)\mathbb{E}(Y \mid Z) = \mathbb{E}\big[(X - \mathbb{E}(X \mid Z))(Y - \mathbb{E}(Y \mid Z)) \mid Z\big]$$

  - **Theorem 17. Law of Total Covariance:**

$$\mathrm{Cov}(X,Y) = \mathbb{E}\big(\mathrm{Cov}(X,Y \mid Z)\big) + \mathrm{Cov}\big(\mathbb{E}(X \mid Z), \mathbb{E}(Y \mid Z)\big)$$

5.1.5 Discrete Random Variable Distributions

Binomial: Binomial(n,p) (sum of n Bernoulli random variables)

- Mass function: $\Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$
- Distribution: $\Pr(X \leq k) = \sum_{i=0}^{k} \binom{n}{i} p^i (1-p)^{n-i}$
- Expectation: $\mathbb{E}(X) = np$
- Variance: $\mathrm{Var}(X) = np(1-p)$

Poisson: Poisson(λ): an approximation of the binomial distribution for n very large, p very small, $np \to \lambda \in (0, \infty)$.

- Mass function:
$$\Pr(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}$$
- Distribution: $\Pr(X \leq k) = \sum_{i=0}^{k} \frac{e^{-\lambda}\lambda^i}{i!}$
- Expectation: $\mathbb{E}(X) = \lambda$ (derive from basic definitions)
- Variance: $\mathrm{Var}(X) = \lambda$

Geometric: $\mathrm{G}_1(p)$: the number of Bernoulli trials before the first success.

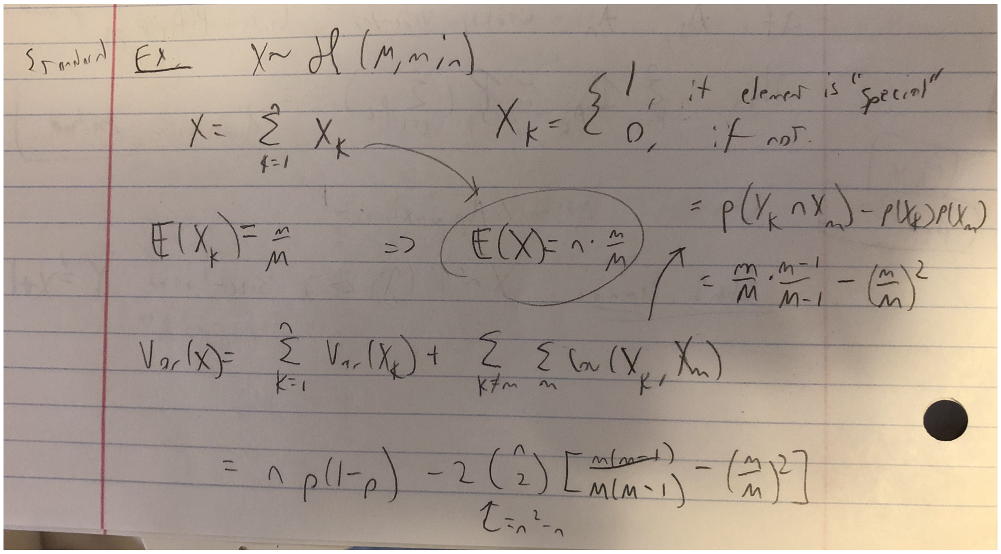
- Mass function: $\Pr(X = k) = p(1-p)^{k-1}$
- Distribution: $\Pr(X \leq k) = \sum_{i=1}^{k} p(1-p)^{k-1}$
- Expectation: $\mathbb{E}(X) = 1/p$
- Variance: $\mathrm{Var}(X) = (1-p)/p^2$

Negative binomial: NB(r, p): The number of Bernoulli trials required for r successes. (Can be derived as the sum of r identically distributed geometric random variables.)

- Mass function: $\Pr(X = k) = \binom{k-1}{r-1} p^r (1-p)^{k-r}$
- Distribution: $\Pr(X \leq k) = \sum_{i=r}^{k} \binom{i-1}{r-1} p^r (1-p)^{i-r}$
- Expectation: $\mathbb{E}(X) =$
- Variance: $\mathrm{Var}(X) =$

Hypergeometric: Hypergeometric(N, M, K): When drawing a sample of size K from a group of N items, M of which are special, the number of special items retrieved.

- Mass function:
$$\Pr(X = k) = \frac{\binom{M}{k}\binom{N-M}{K-k}}{\binom{N}{K}}$$
- Distribution:
$$\Pr(X \leq k) = \sum_{i=0}^{k} \frac{\binom{M}{i}\binom{N-M}{K-i}}{\binom{N}{K}}$$
- Expectation: $\mathbb{E}(X) =$ (find by indicator method)



- Variance: $\mathrm{Var}(X) =$ (find by indicator method)

5.1.6 Indicator Method

Proposition 18. If $\mathbf{1}_{A_k}$ is an indicator then

(a)
$$\mathrm{Cov}(\mathbf{1}_{A_k}, \mathbf{1}_{A_m}) = \mathbb{E}(\mathbf{1}_{A_k}\mathbf{1}_{A_m}) - \mathbb{E}(\mathbf{1}_{A_k})\mathbb{E}(\mathbf{1}_{A_m}) = \Pr(A_k \cap A_m) - \Pr(A_k)\Pr(A_m)$$

(b)
$$\mathrm{Var}(\mathbf{1}_{A_k}) = \mathbb{E}(\mathbf{1}_{A_k}^2) = \mathbb{E}(\mathbf{1}_{A_k})^2 = \Pr(A_k) - (\Pr(A_k))^2$$

Theorem 19. *X is independent of Y if and only if X is independent of $\mathbf{1}_A$, $A \in Y$.*

Example problems: 505A Homework 3 problem 9(a)

Worked examples in p. 56 - 59 of Grimmett and Stirzaker 3rd edition.

5.1.7 Linear transformations of random variables

5.1.8 Poisson Paradigm (Poisson approximation for indicator method)

Theorem 20. (Theorem 4.12.9, p. 129 of Grimmett and Stirzaker.) *Let A_i be an event. If $X = \sum_{i=1}^{m} \mathbf{1}_{A_i}$ where $\mathbf{1}_{A_i}$ is an indicator variable for A_i, and the A_i are only weakly dependent on each other, then*

$$\text{As } m \to \infty, \quad X \sim \mathrm{Poisson}(\mathbb{E}(X))$$

More specifically, let B_i be n independent Bernoulli random variables with probabilities p_i. If $Y = \sum_{i=1}^{n} B_i$ then

$$\text{As } n \to \infty, \quad Y \sim \text{Poisson}\left(\mathbb{E}\Big(\sum_i B_i\Big)\right) = \text{Poisson}\left(\sum_i \mathbb{E}B_i\right) = \text{Poisson}\left(\sum_i p_i\right)$$

5.1.9 Asymptotic Distributions

Proposition 21.

$$e^x = \lim_{n\to\infty}\left(1+\frac{x}{n}\right)^n$$

Theorem 22. Stirling's Formula:

$$n! \approx n^n e^{-n}\sqrt{2\pi n}$$

5.2 Worked problems

5.2.1 Example Problems That Will Likely Appear on Midterm

Fall 2011 Problem 1 (same as HW1 problem 5; similar to HW3 problem 2(5); likely to be question 1 on the midterm.) True or false: if A and B are events such that $0 < \Pr(A) < 1$ and $\Pr(B \mid A) = \Pr(B \mid A^c)$, then A and B are independent.

Solution. A and B are independent if and only if

$$\Pr(A \cap B) = \Pr(A) \cdot \Pr(B)$$

We know that

$$\Pr(B) = \Pr(B|A) \cdot \Pr(A) + \Pr(B|A^c) \cdot \Pr(A^c)$$

$$= \Pr(B|A) \cdot \Pr(A) + \Pr(B|A) \cdot (1 - \Pr(A)) = \Pr(B|A) \cdot \Pr(A) + \Pr(B|A) - \Pr(B|A) \cdot \Pr(A)$$

$$= \Pr(B|A)$$

Also, we know that since $\Pr(A) \neq 0$,

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

Per above $\Pr(B|A) = \Pr(B)$, so we have

$$\Pr(B) = \frac{\Pr(A \cap B)}{\Pr(A)}$$

$$\Pr(A \cap B) = \Pr(A) \cdot \Pr(B)$$

which is what we were trying to prove. So the answer is $\boxed{\text{true.}}$

Similar problem: HW3 Problem 2(5). Verify: $\mathbb{E}(X \mid Y) = \mathbb{E}(X)$ if X and Y are independent.

Solution. X and Y are independent if and only if

$$\Pr(X \cap Y) = \Pr(X) \cdot \Pr(Y) \iff \Pr(X = x \cap Y = y) = \Pr(X = x)\Pr(Y = y)$$

$$\iff \Pr(X = x \mid Y = y) \cdot \Pr(Y = y) = \Pr(X = x)\Pr(Y = y) \iff \Pr(X = x \mid Y = y) = \Pr(X = x)$$

$$\implies E(X \mid Y) = \sum_x x \cdot \Pr(X = x \mid Y = y) = \sum_x x \cdot \Pr(X = x) = \mathbb{E}(X)$$

Fall 2014 Problem 1 (likely to be question 2 on the midterm). Let A and B be two events with $0 < \Pr(A) < 1,\ 0 < \Pr(B) < 1$. Define the random variables $\xi = \xi(\omega)$ and $\eta = \eta(\omega)$ by

$$\xi(\omega) = \begin{cases} 5 & \text{if } \omega \in A \\ -7 & \text{if } \omega \notin A \end{cases}, \qquad \eta(\omega) = \begin{cases} 2 & \text{if } \omega \in B \\ 3 & \text{if } \omega \notin B \end{cases}$$

True or false: the events A and B are independent if and only if the random variables ξ and η are uncorrelated?

Solution. ($\implies$) Suppose A and B are independent. Then ξ and η are uncorrelated if and only if $\mathbb{E}(\xi\eta) = \mathbb{E}(\xi)\mathbb{E}(\eta)$. We can write $\xi = 5 \cdot \mathbf{1}_A - 7 \cdot \mathbf{1}_{A^c}$ and $\eta = 2 \cdot \mathbf{1}_B + 3 \cdot \mathbf{1}_{B^c}$. So we have

$$\xi\eta = (5 \cdot \mathbf{1}_A - 7 \cdot \mathbf{1}_{A^c})(2 \cdot \mathbf{1}_B + 3 \cdot \mathbf{1}_{B^c}) = 10 \cdot \mathbf{1}_{A \cap B} + 15 \cdot \mathbf{1}_{A \cap B^c} - 14 \cdot \mathbf{1}_{A^c \cap B} - 21 \cdot \mathbf{1}_{A^c \cap B^c}$$

$$\implies \mathbb{E}(\xi\eta) = 10\Pr(A \cap B) + 15\Pr(A \cap B^c) - 14\Pr(A^c \cap B) - 21\Pr(A^c \cap B^c)$$

Then

$$\mathbb{E}(\xi)\mathbb{E}(\eta) = (5\Pr(A) - 7\Pr(A^c))(2\Pr(B) + 3\Pr(B^c))$$

$$= 10\Pr(A \cap B) + 15\Pr(A \cap B^c) - 14\Pr(A^c \cap B) - 21\Pr(A^c \cap B^c) = \mathbb{E}(\xi\eta)$$

where the second-to-last step follows from the independence of A and B. Therefore η and ξ are uncorrelated.

($\Longleftarrow$) Now suppose η and ξ are uncorrelated. Then ξ and η are independent if and only if $\Pr(\xi \cap \eta) = \Pr(\xi)\Pr(\eta)$. Define

$$\alpha(\omega) = \xi(\omega) + 7 = \begin{cases} 12 & \text{if } \omega \in A \\ 0 & \text{if } \omega \notin A \end{cases}, \qquad \beta(\omega) = \eta(\omega) - 3 = \begin{cases} -1 & \text{if } \omega \in B \\ 0 & \text{if } \omega \notin B \end{cases}$$

Then we have

$$(\alpha\beta)(\omega) = \begin{cases} -12 & \text{if } \omega \in A \cap B \\ 0 & \text{otherwise} \end{cases}$$

Then

$$\mathbb{E}(\xi\eta) = \mathbb{E}\big[(\alpha - 7)(\beta + 3)\big] = \mathbb{E}(\alpha\beta) + 3\mathbb{E}(\alpha) - 7\mathbb{E}(\beta) - 21$$

$$\mathbb{E}(\xi)\mathbb{E}(\eta) = (\mathbb{E}(\alpha) - 7)(\mathbb{E}(\beta) + 3) = \mathbb{E}(\alpha)\mathbb{E}(\beta) - 7\mathbb{E}(\beta) + 3\mathbb{E}(\alpha) - 21$$

Since by assumption $\mathbb{E}(\xi\eta) = \mathbb{E}(\xi)\mathbb{E}(\eta)$, this yields $\mathbb{E}(\alpha\beta) = \mathbb{E}(\alpha)\mathbb{E}(\beta)$. But

$$\mathbb{E}(\alpha\beta) = -12\Pr(A \cap B), \qquad \mathbb{E}(\alpha)\mathbb{E}(\beta) = 12\Pr(A)(-1)\Pr(B) = -12\Pr(A)\Pr(B)$$

Therefore $\Pr(\xi \cap \eta) = \Pr(\xi)\Pr(\eta)$ and ξ and η are independent.

HW1 Problem 8. Two people, A and B, are involved in a duel. The rules are simple: shoot at each other once; if at least one is hit, the duel is over, if both miss, repeat (go to the next round), and so on. Denote by p_A and p_B the probabilities that A hits B and B hits A with one shot, and assume that that hitting/missing is independent from round to round. Compute the probabilities of the following events: (a) the duel ends and A is not hit; (b) the duel ends and both are hit; (c) the duel ends after round number n; (d) the duel ends after round number n GIVEN that A is not hit; (e) the duel ends after n rounds GIVEN that both are hit; (f) the duel goes on forever.

Solution.

(a) Let A_k denote the event that the duel is ended by A shooting B in the kth round (with neither person being shot in the first $k-1$ rounds). Note that $\{A_k | k = 1, 2, \ldots\}$ are all mutually exclusive. Therefore the probability of the duel ending without A being hit is $\sum_{k=1}^{\infty} A_k$. Because the probabilities in each round are constant and independent,

$$A_k = (1 - p_A)^{k-1} p_A (1 - p_B)^k$$

So the probability that the duel ends and A is not hit is

$$\sum_{k=1}^{\infty} A_k = \sum_{k=1}^{\infty} (1 - p_A)^{k-1} p_A (1 - p_B)^k = p_A(1 - p_B) \sum_{k=1}^{\infty} (1 - p_A)^{k-1} (1 - p_B)^{k-1}$$

This is an infinite geometric series. Since the ratio $(1-p_A)(1-p_B)$ has absolute value less than 1, the sum can be calculated.

$$\sum_{k=1}^{\infty} A_k = p_A(1-p_B)\cdot\frac{1}{1-(1-p_A)(1-p_B)} = \frac{p_A(1-p_B)}{p_A+p_B-p_Ap_B} = \boxed{\frac{p_A(1-p_B)}{p_A(1-p_B)+p_B}}$$

(b) Similar to part (a). Let C_k denote the event that the duel is ended with both players being shot in the kth round (with neither person being shot in the first $k-1$ rounds). Again, $\{C_k|k=1,2,\ldots\}$ are all mutually exclusive, so the probability of the duel ending in these circumstances is $\sum_{k=1}^{\infty} C_k$. We have

$$C_k = (1-p_A)^{k-1}p_A(1-p_B)^{k-1}p_B$$

$$\sum_{k=1}^{\infty} C_k = \sum_{k=1}^{\infty}(1-p_A)^{k-1}p_A(1-p_B)^{k-1}p_B = p_Ap_B\sum_{k=1}^{\infty}(1-p_A)^{k-1}(1-p_B)^{k-1}$$

$$= p_Ap_B\cdot\frac{1}{1-(1-p_A)(1-p_B)} = \boxed{\frac{p_Ap_B}{p_A+p_B-p_Ap_B}}$$

Note that this value is less than the answer from part (a) if $p_B < \frac{1}{2}$ and greater if $p_B > \frac{1}{2}$

(c) Let B_k denote the event that the duel is ended by B shooting A in the kth round (with neither person being shot in the first $k-1$ rounds), with

$$B_k = (1-p_A)^k p_B(1-p_B)^{k-1}$$

Let A_k and C_k be defined as above. Note that $\{A_k|k=1,2,\ldots\},\{B_k|k=1,2,\ldots\},\{C_k|k=1,2,\ldots\}$ are all mutually exclusive, and that the event that the duel ends in round n is $\{A_n \cup B_n \cup C_n\}$. So the probability of the duel ending in round n is

$$\Pr(A_n \cup B_n \cup C_n) = \Pr(A_n) + \Pr(B_n) + \Pr(C_n)$$

$$= (1-p_A)^{n-1}p_A(1-p_B)^n + (1-p_A)^n p_B(1-p_B)^{n-1} + (1-p_A)^{n-1}p_A(1-p_B)^{n-1}p_B$$

$$= (1-p_A)^{n-1}(1-p_B)^{n-1}\left[p_A(1-p_B) + (1-p_A)p_B + p_Ap_B\right]$$

$$= \boxed{(1-p_A)^{n-1}(1-p_B)^{n-1}(p_A+p_B-p_Ap_B)}$$

(d) Let A_k, B_k, C_k be defined as above. The event that the duel ends at round n without A being hit is given by $\{A_n\}$.

$$\Pr(A_n) = \boxed{(1-p_A)^{n-1}p_A(1-p_B)^n}$$

(e) Let A_k, B_k, C_k be defined as above. The event that the duel ends at round n with both players being hit is given by $\{C_n\}$.

$$\Pr(C_n) = \boxed{(1-p_A)^{n-1}p_A(1-p_B)^{n-1}p_B}$$

(f) Let A_k, B_k, C_k be defined as above. The probability that the duel never ends is equal to 1 - the probability that the duel ends at some point, which is $\{A_k | k = 1, 2, \ldots\} \cup \{B_k | k = 1, 2, \ldots\} \cup \{C_k | k = 1, 2, \ldots\}$. Since all of these events are mutually exclusive, we have

$$1 - \Pr(\{A_k | k = 1, 2, \ldots\} \cup \{B_k | k = 1, 2, \ldots\} \cup \{C_k | k = 1, 2, \ldots\}) = 1 - \sum_{k=1}^{\infty} \left(A_k + B_k + C_k\right)$$

$$= 1 - \sum_{k=1}^{\infty} \left((1-p_A)^{k-1} p_A (1-p_B)^k + (1-p_A)^k p_B (1-p_B)^{k-1} + (1-p_A)^{k-1} p_A (1-p_B)^{k-1} p_B\right)$$

$$= 1 - \left[p_A(1-p_B) + (1-p_A)p_B + p_A p_B\right] \sum_{k=1}^{\infty} (1-p_A)^{k-1}(1-p_B)^{k-1}$$

$$= 1 - \left[p_A 1 - p_A p_B) + p_B - p_A) p_B + p_A p_B\right] \cdot \frac{1}{1 - (1-p_A)(1-p_B)}$$

$$= 1 - \frac{p_A - p_A p_B + p_B - p_A p_B + p_A p_B}{p_A + p_B - p_A p_B} = 1 - \frac{p_A + p_B - p_A p_B}{p_A + p_B - p_A p_B} = \boxed{0}$$

Similar: HW3 Problem 2 (parts 1 - 4). Verify:

(1) $\mathbb{E}(\mathbb{E}(X \mid Y)) = \mathbb{E}(X)$

(2) $\mathbb{E}(g(Y)X \mid Y) = g(Y)\mathbb{E}(X \mid Y)$

(3) $\mathrm{Cov}(\mathbb{E}(X \mid Y), Y) = \mathrm{Cov}(X, Y)$

(4) Y and $X - \mathbb{E}(X \mid Y)$ are uncorrelated.

Solution.

(1)

$$\mathbb{E}\big(\mathbb{E}(X \mid Y)\big) = \sum_y \mathbb{E}(X \mid Y) \Pr(Y = y) = \sum_y \left[\sum_x x \cdot \Pr(X = x \mid Y = y) \Pr(Y = y)\right]$$

$$= \sum_y \left[\sum_x x \cdot \Pr(X = x \cap Y = y)\right] = \sum_y \left[\sum_x x \cdot \Pr(Y = y \mid X = x) \cdot \Pr(X = x)\right]$$

$$= \sum_x \left[x \cdot \Pr(X = x) \cdot \sum_y \big(\Pr(Y = y \mid X = x)\big)\right] = \sum_x \left[x \cdot \Pr(X = x) \cdot 1\right]$$

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

(2) 2

(3)

$$\text{Cov}(\mathbb{E}(X \mid Y), Y) = \mathbb{E}\Big(\Big[\mathbb{E}(X \mid Y) - \mathbb{E}\big(\mathbb{E}(X \mid Y)\big)\Big]\Big[Y - \mathbb{E}(Y)\Big]\Big)$$

$$= \mathbb{E}\Big(\Big[\mathbb{E}(X \mid Y) - \mathbb{E}(X)\Big]\Big[Y - \mathbb{E}(Y)\Big]\Big) = \mathbb{E}\Big(\mathbb{E}(X \mid Y)Y - \mathbb{E}(X)Y - \mathbb{E}(X \mid Y)\mathbb{E}(Y) + \mathbb{E}(X)\mathbb{E}(Y)\Big)$$

$$= \mathbb{E}\big(\mathbb{E}(X \mid Y)Y\big) - \mathbb{E}(X)\mathbb{E}(Y) - \mathbb{E}(Y)\mathbb{E}\big(\mathbb{E}(X \mid Y)\big) + \mathbb{E}(X)\mathbb{E}(Y) = \mathbb{E}\big(\mathbb{E}(X \mid Y)Y\big) - \mathbb{E}(Y)\mathbb{E}(X)$$

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

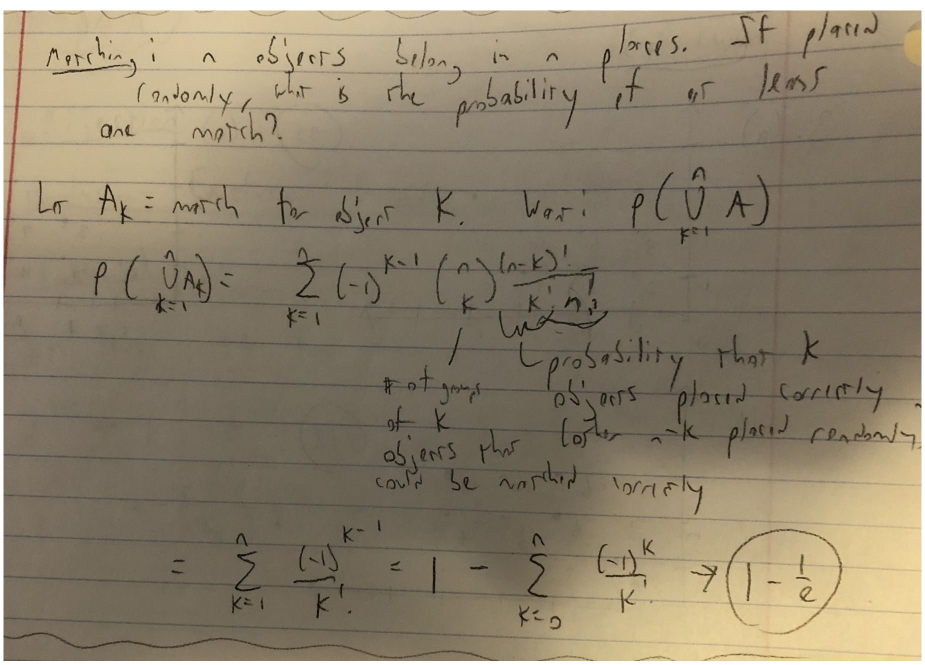
(4) Y and $X - \mathbb{E}(X \mid Y)$ are uncorrelated if and only if $\text{Cov}\big(Y, X - \mathbb{E}(X \mid Y)\big) = 0 \iff \mathbb{E}(Y \cdot \big[X - \mathbb{E}(X \mid Y)\big]) - \mathbb{E}(Y)\mathbb{E}(X - \mathbb{E}(X \mid Y)) = 0$.

$$\mathbb{E}(Y \cdot \big[X - \mathbb{E}(X \mid Y)\big]) - \mathbb{E}(Y)\mathbb{E}(X - \mathbb{E}(X \mid Y)) = \mathbb{E}(YX - Y\mathbb{E}(X \mid Y)) - \mathbb{E}(Y)\mathbb{E}(X) + \mathbb{E}(Y)\mathbb{E}\big(\mathbb{E}(X \mid Y)\big))$$

$$= \mathbb{E}(YX) - \mathbb{E}\big(Y\mathbb{E}(X \mid Y)\big) - \mathbb{E}(Y)\mathbb{E}(X) + \mathbb{E}(Y)\mathbb{E}(X) = \mathbb{E}(YX) - \mathbb{E}(YX) = 0$$

Remaining problems are likely to be indicator method.

5.2.2 Problems we did in class that professor mentioned



Matching: n objects belong in n places. If placed randomly, what is the probability of at least one match?

Let A_k = match for object k. Want: $P(\bigcup_{k=1}^{n} A)$

$$P\Big(\bigcup_{k=1}^{n} A_k\Big) = \sum_{k=1}^{n} (-1)^{k-1} \binom{n}{k} \frac{(n-k)!}{n!}$$

# of groups of k objects that could be matched correctly

probability that k objects placed correctly (other n-k placed randomly)

$$= \sum_{k=1}^{n} \frac{(-1)^{k-1}}{k!} = 1 - \sum_{k=0}^{n} \frac{(-1)^k}{k!} \to 1 - \frac{1}{e}$$

Standard Example 2 Matches (n objects, n places)

$X = \#$ of matches

$X_k = 1$ of objects matched at correct location

$$X = \sum_{k=1}^{n} X_k$$

$$P(X_k = 1) = \frac{1}{n} \qquad \mathbb{E}(X) = 1$$

$$P(X_k = 1, X_m = 1) = \frac{1}{n} \cdot \frac{1}{n-1}$$

$$\mathrm{Var}(X) = \underbrace{n \cdot \frac{1}{n}\left(1 - \frac{1}{n}\right)}_{np(1-p)} + n(n-1)\left(\frac{1}{n} \cdot \frac{1}{n-1} - \frac{1}{n^2}\right)$$

Variance Problem 09/21 If $\mathbb{E}(X \mid Y) = Y, \mathbb{E}(Y \mid X) = X, \mathbb{E}(X^2) < \infty, \mathbb{E}(Y^2) < \infty$, show $\mathbb{E}(X - Y)^2 = 0$ (or equivalently, show $\Pr(X = Y) = 1$).

Solution.

$$\mathbb{E}(X - Y)^2 = \mathbb{E}(X^2 - 2XY + Y^2) = \mathbb{E}(X^2) - 2\mathbb{E}(XY) + \mathbb{E}(Y^2)$$

$$\mathbb{E}(XY) = \mathbb{E}(\mathbb{E}(XY \mid Y)) = \mathbb{E}(Y\mathbb{E}(X \mid Y)) = \mathbb{E}(Y \cdot Y) = \mathbb{E}(Y^2)$$

Also,

$$\mathbb{E}(XY) = \mathbb{E}((XY \mid X)) = \mathbb{E}(X\mathbb{E}(Y \mid X)) = \mathbb{E}(X \cdot X) = \mathbb{E}(X^2)$$

Therefore

$$\mathbb{E}(X - Y)^2 = 0$$

Spring 2018 Problem 2 (did not complete)

2. Consider positions 1 to n arranged in a circle, so that 2 comes after 1, 3 comes after 2, ..., n comes after $n-1$, and 1 comes after n. Similarly, take 1 to n as values, with cyclic order, and consider all $n!$ ways to assign values to positions, bijectively, with all $n!$ possibilities equally likely. For $i = 1$ to n, let X_i be the indicator that position i and the one following are filled in with two consecutive values in increasing order, and define

$$S_n = \sum_{i=1}^{n} X_i, \quad T_n = \sum_{i=1}^{n} iX_i$$

For example, with $n = 6$ and the the circular arrangement 314562, we get $X_3 = 1$ since 45 are consecutive in increasing order, and similarly $X_4 = X_6 = 1$, so that $S_6 = 3, T_6 = 13$.

a) Compute the mean and the variance of S_n.

b) Compute the mean and the variance of T_n.

Fall 2008 Problem 2 (HW1 Problem 10). Consider a lottery with n^2, tickets, of which only n tickets win prizes. Let p_n be the probability that, out of n randomly selected tickets, at least one wins a prize. Compute $\lim_{n\to\infty} p_n$.

Solution. There are $\binom{n^2}{n}$ possible sets of n tickets. The number of these sets that do not contain at least one winner (that is, they only contain members of the $n^2 - n$ losing tickets) is $\binom{n^2-n}{n}$. Therefore the probability of selecting a set of n tickets that contains at least one winner is

$$p_n = 1 - \binom{n^2-n}{n} \Big/ \binom{n^2}{n} = 1 - \frac{(n^2-n)!}{n!(n^2-n-n)!} \Big/ \frac{(n^2)!}{(n^2-n)!n!} = 1 - \frac{(n^2-n)!}{n!(n^2-2n)!} \cdot \frac{(n^2-n)!n!}{(n^2)!}$$

$$= 1 - \frac{(n^2-n)!}{(n^2-2n)!} \cdot \frac{(n^2-n)!}{(n^2)!} = 1 - \prod_{i=0}^{n-1}(n^2-n-i) \Big/ \prod_{i=0}^{n-1}(n^2-i) = 1 - \prod_{i=0}^{n-1} \frac{n^2-n-i}{n^2-i}$$

$$= 1 - \prod_{i=0}^{n-1} \left(\frac{n^2-i}{n^2-i} - \frac{n}{n^2-i} \right) = 1 - \prod_{i=0}^{n-1} \left(1 - \frac{n}{n^2-i} \right)$$

Therefore

$$\lim_{n\to\infty} p_n = \lim_{n\to\infty} \left[1 - \prod_{i=0}^{n-1} \left(1 - \frac{n}{n^2-i} \right) \right] = 1 - \lim_{n\to\infty} \prod_{i=0}^{n} \left(1 - \frac{n}{n^2-i} \right) = 1 - \lim_{n\to\infty} \prod_{i=0}^{n} \left(1 - \frac{n \cdot \frac{1}{n}}{\frac{n^2}{n} - \frac{i}{n}} \right)$$

$$= 1 - \lim_{n\to\infty} \prod_{i=0}^{n} \left(1 - \frac{1}{n - \frac{i}{n}} \right) = 1 - \lim_{n\to\infty} \prod_{i=0}^{n} \left(1 - \frac{1}{n} \right) = 1 - \lim_{n\to\infty} \left(1 - \frac{1}{n} \right)^n = \boxed{1 - \exp(-1)}$$

5.2.3 Problems we did on homework

Fall 2017 Problem 3 (HW3 Problem 8—almost full solution)

Problem 8. Let $U_1, U_2, \ldots$ be iid random variables, uniformly distributed on $[0,1]$, and let N be a Poisson random variable with mean value equal to one. Assume that N is independent of $U_1, U_2, \ldots$ and define

$$Y = \begin{cases} 0, & \text{if } N = 0, \\ \max\limits_{1 \le i \le N} U_i, & \text{if } N > 0. \end{cases}$$

Compute the expected value of Y.

Solution.

Since Y is a function of N, let $Y = y(N)$. By the Law of the Unconscious Statistician (Theorem 11),

$$\mathbb{E}(Y) = \mathbb{E}\big(\mathbb{E}(Y \mid N)\big) = \mathbb{E}\big(\mathbb{E}(\max_{1 \le i \le N} U_i \mid N = n)\big)$$

Let $Z_n = \max_{1 \le i \le n} U_i$. The cdf of Z_n can be calculated as follows:

$$\Pr(Z_n \le x) = \Pr(\max_{1 \le i \le n} U_i \le x) = \Pr(U_1 \le x \cap U_2 \le x \cap \ldots \cap U_n \le x) = x^n$$

for $x \in [0,1]$. Therefore the pdf of Z_n is its derivative, nx^{n-1}. So we have

$$\mathbb{E}(\max_{1 \le i \le N} U_i \mid N = n) = \mathbb{E}(Z_n) = \int_0^1 x n x^{n-1} dx = n \int_0^1 x^n dx = n \frac{x^{n+1}}{n+1}\bigg|_0^1 = \frac{n}{n+1}$$

Plugging this into the expression for $\mathbb{E}(Y)$ yields

$$\mathbb{E}(Y) = \mathbb{E}\left(\frac{N}{N+1}\right) = \sum_{n=1}^{\infty} \frac{n}{n+1} \Pr(N = n) = \sum_{n=1}^{\infty} \frac{n}{n+1} \frac{\exp(-1) 1^n}{n!} = \boxed{\frac{1}{e} \sum_{n=1}^{\infty} \frac{n}{(n+1)!}}$$

Fall 2013 Problem 3/Spring 2011 Problem 2 (HW3 Problem 9; coupon collector problem) Only parts I didn't do: Let D be the event that no box receives more than 1 ball. Fix $a \in (0,1)$. If both $n, d \to \infty$ together, what relation must they satisfy in order to have $\Pr(D) \to a$?

HW3 Problem 9. Consider n (different) balls placed at random in m boxes so that each of m^n configurations is equally likely.

(a) Compute the expected value and the variance of the number of empty boxes.

(b) Show that if $\lim_{m,n\to\infty} m \exp(-n/m) = \lambda \in (0, \infty)$, then, in the same limit, the number of empty boxes has Poisson distribution with parameter λ.

(c) For $k \ge 1$ such that $k + 3 \le m$, define the event A_k that the boxes $k, k+1, k+2, k+3$ are empty. Assuming that $m > 8$, compute $\Pr(A_1 \cup A_3 \cup A_5)$. How will the answer change if $m = 8$?

(d) Now imagine that the balls are dropped one-by-one (with each ball equally likely to go into any of the m boxes, independent of all other balls), and denote by N_m the minimal number of balls required to fill all the boxes. Compute $\mathbb{E}(N_m), \mathrm{Var}(N_m)$ and

$$\lim_{m\to\infty} \Pr\left(\frac{N_m - m\log m}{m} \le x\right)$$

Solution.

(a) Let A_i be the event that the ith box is empty. Let $\mathbf{1}_{A_i}$ be the indicator for A_i. Then $X = \sum_{i=1}^m \mathbf{1}_{A_i}$.

$$\mathbb{E}(X) = \mathbb{E}\left(\sum_{i=1}^m \mathbf{1}_{A_i}\right) = \sum_{i=1}^m \left(\mathbb{E}\mathbf{1}_{A_i}\right) = \sum_{i=1}^m \Pr(A_i) = \sum_{i=1}^m \left(\frac{m-1}{m}\right)^n = \boxed{\frac{(m-1)^n}{m^{n-1}}}$$

$$\mathrm{Var}(X) = \mathrm{Var}\left(\sum_{i=1}^m \mathbf{1}_{A_i}\right) = \sum_{i=1}^m \mathrm{Var}(\mathbf{1}_{A_i}) + 2\sum_{1\le i<j\le m} \mathrm{Cov}(\mathbf{1}_{A_i}, \mathbf{1}_{A_j})$$

$$\mathrm{Var}(\mathbf{1}_{A_i}, \mathbf{1}_{A_j}) = \mathbb{E}(\mathbf{1}_{A_i}\mathbf{1}_{A_i}) - \mathbb{E}(\mathbf{1}_{A_i})^2 = \Pr(A_i \cap A_i) - \Pr(A_i)^2 = \left(\frac{m-1}{m}\right)^n - \left(\frac{m-1}{m}\right)^{2n}$$

$$\mathrm{Cov}(\mathbf{1}_{A_i}, \mathbf{1}_{A_j}) = \mathbb{E}(\mathbf{1}_{A_i}\mathbf{1}_{A_j}) - \mathbb{E}(\mathbf{1}_{A_i})\mathbb{E}(\mathbf{1}_{A_j}) = \Pr(A_i \cap A_j) - \Pr(A_i)\Pr(A_j) = \left(\frac{m-2}{m}\right)^n - \left(\frac{m-1}{m}\right)^{2n}$$

$$\implies \mathrm{Var}(X) = m \cdot \left[\left(\frac{m-1}{m}\right)^n - \left(\frac{m-1}{m}\right)^{2n}\right] + \frac{m!}{(m-2)!}\left[\left(\frac{m-2}{m}\right)^n - \left(\frac{m-1}{m}\right)^{2n}\right]$$

$$= \frac{(m-1)^n}{m^{n-1}} - \frac{(m-1)^{2n}}{m^{2n-1}} + (m^2 - m)\left[\left(\frac{m-2}{m}\right)^n - \left(\frac{m-1}{m}\right)^{2n}\right]$$

$$\boxed{\mathrm{Var}(X) = \frac{(m-1)^n}{m^{n-1}} - \frac{(m-1)^{2n}}{m^{2n-1}} + (m-1)\left[\frac{(m-2)^n}{m^{n-1}} - \frac{(m-1)^{2n}}{m^{2n-1}}\right]}$$

(b) Note that

$$X = \sum_{i=1}^m \mathbf{1}_{A_i}$$

and that the A_i are only weakly dependent on each other, especially as m and n increase. Therefore as $m, n \to \infty$, the Poisson paradigm suggests $X \sim \mathrm{Poisson}(\mathbb{E}(X))$. We have

$$\mathbb{E}(X) = \frac{(m-1)^n}{m^{n-1}}$$

so

$$\lim_{n,m\to\infty} \mathbb{E}(X) = \lim_{n,m\to\infty} m \cdot \left(\frac{m-1}{m}\right)^n = \lim_{n,m\to\infty} m \cdot \left(1 - \frac{1}{m}\right)^n = \lim_{n,m\to\infty} m \cdot \left[\left(1 - \frac{1}{m}\right)^m\right]^{n/m}$$

$$\approx \lim_{n,m\to\infty} m \cdot \left[e^{-1}\right]^{n/m} = \lim_{n,m\to\infty} m e^{-n/m}$$

Using

$$\lim_{m,n\to\infty} m\exp(-n/m) = \lambda \in (0,\infty)$$

we have $\boxed{X \sim \text{Poisson}(\lambda) \text{ as } m,n\to\infty}$.

(c)

$$\Pr(A_1\cup A_3\cup A_5) = \Pr(A_1)+\Pr(A_3)+\Pr(A_5)-\Pr(A_1\cap A_3)-\Pr(A_1\cap A_5)-\Pr(A_3\cap A_5)+\Pr(A_1\cap A_3\cap A_5)$$

We have

$$\Pr(A_1) = \Pr(A_3) = \Pr(A_5) = \left(\frac{m-4}{m}\right)^n$$

$$\Pr(A_1\cap A_3) = \Pr(A_3\cap A_5) = \left(\frac{m-6}{m}\right)^n$$

$$\Pr(A_1\cap A_5) = \Pr(A_1\cap A_3\cap A_5) = \left(\frac{m-8}{m}\right)^n$$

Therefore

$$\Pr(A_1\cup A_3\cup A_5) = 3\left(\frac{m-4}{m}\right)^n - 2\left(\frac{m-6}{m}\right)^n = \boxed{\frac{3(m-4)^n - 2(m-6)^n}{m^n}}$$

(d) N_m is the minimal number of balls required to fill all the boxes. Let T_i be the number of balls that have to be dropped to fill the ith box after $i-1$ boxes have been filled. The probability of filling a new box after $i-1$ boxes have been filled is $\frac{m-(i-1)}{m}$. Therefore T_i has a geometric distribution with $E(T_i) = \frac{m}{m-(i-1)}$. Since $N_m = \sum_{i=1}^m T_i$, we have

$$\mathbb{E}(N_m) = \mathbb{E}\left(\sum_{i=1}^m T_i\right) = \sum_{i=1}^m \mathbb{E}(T_i) = \sum_{i=1}^m \frac{m}{m-(i-1)} = \boxed{m\sum_{i=1}^m \frac{1}{i}}$$

Because the T_i are independent, we have

$$\begin{aligned}\text{Var}(N_m) = \text{Var}\left(\sum_{i=1}^m T_i\right) &= \sum_{i=1}^m \text{Var}(T_i) = \sum_{i=1}^m \left(1-\frac{m-(i-1)}{m}\right) \Bigg/ \left(\frac{m-(i-1)}{m}\right)^2 \\ &= \sum_{i=1}^m \frac{i-1}{m}\cdot\left(\frac{m}{m-(i-1)}\right)^2 = \boxed{m\sum_{i=1}^m \frac{i-1}{[m-(i-1)]^2}}\end{aligned}$$

Finally, to find

$$\lim_{m\to\infty} \Pr\left(\frac{N_m - m\log m}{m} \le x\right)$$

begin by noting that we can also express N_m as

$$\Pr(N_m \le k) = \Pr\left(X_{m,k} = 0\right)$$

where $X_{m,k}$ is defined as X is in part (b) with k being the number of balls that have been dropped so far, $k \in \mathbb{N} \geq m$. (For $k < m$, $\Pr(N_m \leq k) = 0$.)

Again, let $A_{i,k}$ be the event that the ith box is empty after dropping k balls. Then because $X_{m,k} = \sum_{i=1}^{m} \mathbf{1}_{A_{i,k}}$ and the $A_{i,k}$ are only weakly dependent on each other (especially as m becomes large), the Poisson paradigm again suggests that as $m \to \infty$, $X_{m,k} \sim \text{Poisson}(\lambda_k)$ where $\lambda_k = \mathbb{E}(X_{m,k})$ is defined as above. Therefore we have

$$\lim_{m \to \infty} \Pr\left(\frac{N_m - m\log m}{m} \leq x\right) = \lim_{m \to \infty} \Pr\big(N_m \leq xm + m\log m\big) = \lim_{m \to \infty} \Pr\big(X_{m,xm+m\log m}$$

$$= 0) \approx \frac{\exp(-\lambda_{xm+m\log m}) \cdot \lambda_{xm+m\log m}^0}{0!} = \exp\big(-\lambda_{xm+m\log m}\big)$$

And we have

$$\lambda_{xm+m\log m} = \lim_{m \to \infty} m\exp\left(-\frac{xm + m\log m}{m}\right) = \lim_{m \to \infty} m\exp\big(-x - \log m\big) = \lim_{m \to \infty} m/m\exp(-x)$$

$$= \exp(-x)$$

which yields

$$\boxed{\lim_{m \to \infty} \Pr\left(\frac{N_m - m\log m}{m} \leq x\right) = \exp(\exp(-x))}$$

Fall 2012 Problem 1 (HW2 Problem 10/HW 1 Problem 9) Only part I didn't do: Find the mean and variance of $S_n = X_1 + \ldots + X_n$, the total number of white balls added to the urn up to time n.

HW1 Problem 9. An urn contains b black and w white balls. At each step, a ball is removed from the urn at random and then put back together with one more ball of the same color. Compute the probability p_n to get a black ball on step $n, n \geq 1$.

Solution. Step 1:

$$p_1 = \frac{b}{b+w}$$

Step 2: We need to separately consider the cases where a black ball was selected on step 1 (with probability p_1) or a white ball (with probability $1 - p_1$).

$$p_2 = p_1 \cdot \frac{b+1}{b+w+1} + (1-p_1) \cdot \frac{b}{b+w+1} = p_1\left(\frac{b+1}{b+w+1} - \frac{b}{b+w+1}\right) + \frac{b}{b+w+1}$$

$$= p_1\left(\frac{1}{b+w+1} + \frac{1}{p_1}\frac{b}{b+w+1}\right) = p_1\left(\frac{1}{b+w+1} + \frac{b+w}{b}\frac{b}{b+w+1}\right)$$

$$= p_1 \left(\frac{b+w+1}{b+w+1} \right) = p_1$$

$$\implies p_2 = p_1 = \frac{b}{b+w}$$

Step 3: Regardless of the previous steps, there are now $b+w+2$ balls in the urn. Since we know that $p_1 = p_2$, the probability that we have selected k black balls so far (and thus, the probability that there are currently $b+k$ black balls in the urn) is given by

$$\Pr(k \text{ balls chosen in first 2 rounds}) = \binom{2}{k} p_1^k (1-p_1)^{2-k} = \binom{2}{k} \left(\frac{b}{b+w} \right)^k \left(\frac{w}{b+w} \right)^{2-k}$$

$$= \binom{2}{k} \frac{b^k w^{2-k}}{(b+w)^2}$$

for $k \in \{0, 1, 2\}$. Given that we have selected k black balls so far, the probability of selecting a black ball this time is $\frac{b+k}{b+w+2}$. Therefore the probability of selecting a black ball this round is

$$p_3 = \sum_{k=0}^{2} \binom{2}{k} \frac{b^k w^{2-k}}{(b+w)^2} \frac{b+k}{b+w+2} = \frac{1}{(b+w+2)(b+w)^2} \sum_{k=0}^{2} \binom{2}{k} (b+k) b^k w^{2-k}$$

$$= \frac{1}{(b+w+2)(b+w)^2} \left(\binom{2}{0} bw^2 + \binom{2}{1} (b+1)bw + \binom{2}{2} (b+2)b^2 \right)$$

$$= \frac{bw^2 + 2(b+1)bw + (b+2)b^2}{(b+w+2)(b+w)^2} = \frac{b}{b+w} \left(\frac{w^2 + 2bw + 2w + b^2 + 2b}{b^2 + bw + 2b + wb + w^2 + 2w} \right)$$

$$= \frac{b}{b+w} \left(\frac{w^2 + 2bw + 2w + b^2 + 2b}{b^2 + 2bw + 2b + w^2 + 2w} \right) = \frac{b}{b+w} = p_1$$

There seems to be a clear pattern here. Let's find the general formula by induction.

Step $n+1$: Assume that the probability of choosing a black ball on steps $1, 2, \ldots, n$ was $\frac{b}{b+w}$ each time.

(a bunch of boring stuff, then it worked.)

HW2 Problem 10. Random variables $(X_1, \ldots, X_n)$ are called *exchangeable* if $\Pr(X_1 = x_1, \ldots, X_n = x_n) = \Pr(X_{\tau(1)} = x_1, \ldots, X_{\tau(n)} = x_n)$ for all real numbers $x_1, \ldots, x_n$ and every permutation τ of the set $\{1, \ldots, n\}$. In the setting of Problem 9 from Homework 1, let $X_k = 1$ if a white ball is drawn on step k, and $X_k = 0$ otherwise. Show that the random variables $X_1, \ldots, X_n$ are exchangeable for every $n \geq 2$.

Solution. For $n = 2$: There are two cases which we must show are equal to show exchangeability:

$$\Pr(X_1 = 0, X_2 = 1) = \Pr(X_1 = 1, X_2 = 0)$$

First,

$$\Pr(X_1 = 0, X_2 = 1) = \Pr(\text{black first}) \Pr(\text{white second} \mid \text{black first}) = \left(\frac{b}{b+w}\right)\left(\frac{w}{b+w+1}\right)$$

$$\left(\frac{w}{b+w}\right)\left(\frac{b}{b+w+1}\right) = \Pr(X_1 = 1, X_2 = 0)$$

which proves exchangeability for $n = 2$. In the general case, we seek to show that $X_1, \ldots, X_n$ are exchangeable. That is, in all $n+1$ unordered sets $\mathbb{X}_k = \{x_{1k}, x_{2k}, \ldots, x_{nk} \mid x_{ik} \in \{0,1\}, \sum_i x_{ik} = k\}$, in all $\binom{n}{k}$ permutations of $\mathbb{X}_k$,

$$\Pr(\mathbb{X}_{kj} = \Pr(\mathbb{X}_{kj'}$$

where j and j' denote different permutations of $\mathbb{X}_k$. That is,

$$\Pr(X_1 = x_{1k}, X_2 = x_{2k}, \ldots, X_n = x_{nk}) = \Pr(X_{j_1} = x_{1k}, X_{j_2} = x_{2k}, \ldots, X_{j_n} = x_{nk})$$

where $j_1, j_2, \ldots, j_n$ index the permuted variables. Consider $\mathbb{X}_{kj^*}$ where all k white balls are chosen first and all $n-k$ black balls are chosen last. We have

$$\Pr(\mathbb{X}_{kj^*}) = \prod_{i=1}^{k}\left(\frac{w+i-1}{b+w+i-1}\right) \cdot \prod_{i=k+1}^{n}\left(\frac{b+i-k-1}{b+w+i-1}\right)$$

$$= \prod_{i=1}^{n}\left(\frac{1}{b+w+i-1}\right) \cdot \left[\prod_{i=1}^{k}(w+i-1) \prod_{i=k+1}^{n}(b+i-k-1)\right] = \prod_{i=1}^{n}\left(\frac{1}{b+w+i-1}\right) \cdot \left[\prod_{i=1}^{k}(w+i-1) \prod_{i'=1}^{n-k}(b+i'-1)\right]$$

It is easy to see that the leftmost product will always equal the product of the denominators, regardless of the permutation, since one ball is added to the urn after every draw. Similarly, regardless of permutation, the numerator of the probability of drawing the ith white ball will always equal $w + i - 1$, the number of white balls already in the urn. Likewise, the numerator of the probability of drawing the i'th black ball is always $b + i' - 1$. Because multiplication is commutative, all permutations of these numbers will have equal products. Therefore $\Pr(\mathbb{X}_{kj^*}) = \Pr(\mathbb{X}_{kj}$ for all k. That is,

$$\Pr(X_1 = x_1, \ldots, X_n = x_n) = \Pr(X_{\tau(1)} = x_1, \ldots, X_{\tau(n)} = x_n)$$

for all $(x_1, \ldots, x_n) \in \mathbb{R}^n$, all $n \in \mathbb{Z}$ such that $n \geq 2$, all permutations τ.

Homework 2 Problem 2. Consider the function

$$f(x) = \begin{cases} C(2x - x^2) & 0 < x < 2 \\ 0 & \text{otherwise} \end{cases}$$

(a) Could f be a distribution function? If so, determine C.

(b) Could f be a probability density function? If so, determine C.

Solution.

(a) If f is a distribution function, $\lim_{x\to-\infty} f(x) = 0$, $\lim_{x\to\infty} f(x) = 1$, and $f'(x) \geq 0 \; \forall \; x \in \mathbb{R}$. f clearly does not meet the second or third conditions and is therefore $\boxed{\text{not a distribution function.}}$

(b) If f is a density function then $\int_{-\infty}^{\infty} f(x)dx = 1$ and $f(x) \geq 0 \; \forall x \in \mathbb{R}$.

$$\int_{-\infty}^{\infty} f(x)dx = \int_0^2 C(2x - x^2)dx = C\left[x^2 - \frac{x^3}{3}\right]_0^2 = C\left(4 - \frac{8}{3} - 0\right) = C \cdot \frac{4}{3}$$

$$= 1 \iff C = \frac{3}{4}$$

Next we check that f is always nonnegative. It equals zero except on $(0, 2)$.

$$\frac{3}{4}(2x - x^2) \geq 0 \iff x(2 - x) \geq 0 \iff x \in (0, 2)$$

Therefore f is nonnegative $\forall x \in \mathbb{R}$, so $\boxed{f \text{ is a probability density function if } C = \frac{3}{4}.}$

Homework 1 Problem 1.

(I) Seven different gifts are distributed among 10 children. How many different outcomes are possible if every child can receive (a) at most one gift, (b) at most two gifts, (c) any number of gifts?

(II) Answer the same questions if the gifts are identical (but the children are still different).

Solution.

(I) (a) $\binom{10}{7}7! = \boxed{604,800}$

(b) Clearly all outcomes that satisfy part (I)(a) also satisfy these conditions, so we start with $\binom{10}{7}7! = 604,800$ possible outcomes. In addition, the following outcomes are possible:

(i) **A set of 6 children receive gifts; one child receives two gifts.** There are $\binom{10}{6}$ ways to pick a group of 6 children to receive the gifts. Next, there are $\binom{6}{1} = 6$ ways to choose which child receives two gifts. Finally, there are $7!/2!$ unique ways to distribute the gifts among the children once a particular partition is chosen (since order matters for all of the gifts except for the two that are received by the same child).

(ii) **A set of 5 children receive gifts; two children receive two gifts.** There are $\binom{10}{5}$ ways to pick a group of 5 children to receive the gifts. Next, there are $\binom{5}{2}$ ways to choose which of these children receive one gift and which receive two. Finally, there are $7!/(2!2!)$ unique ways to distribute the gifts among the children once a particular partition is chosen (since order matters for all of the gifts except for the two batches of two gifts that are received by the same child).
(Note that without the restriction that a child can receive at most two gifts, another possibility is that 1 child could receive 3 gifts, but that wouldn't work in this case.)

(iii) **A set of 4 children receive gifts; three children each receive two gifts.** There are $\binom{10}{4}$ ways to pick a group of 4 children to receive the gifts. Next, there are $\binom{4}{3} = 4$ ways to choose which of these children receive one gift and which receive two. Finally, there are $7!/(2!2!2!)$ unique ways to distribute the gifts among the children once a particular partition is chosen (since order matters for all of the gifts except for the three batches of two gifts that are received by the same child).
(Again, there are other possibilities for 4 children to receive 7 gifts, but none that satisfy the condition that no child receives more than 2 gifts.)

Clearly each of these outcomes are mutually exclusive. Therefore the answer is

$$\binom{10}{7}7! + \binom{10}{6} \cdot \binom{6}{1} \cdot \frac{7!}{2!} + \binom{10}{5} \cdot \binom{5}{2} \cdot \frac{7!}{2!2!} + \binom{10}{4} \cdot \binom{4}{3} \cdot \frac{7!}{2!2!2!}$$

$$= 7! \cdot \left(\frac{10!}{3!} + \frac{10!}{6!4!} \cdot 6 \cdot \frac{1}{2} + \frac{10!}{5!5!} \cdot \frac{5!}{3!2!} \cdot \frac{1}{4} + \frac{10!}{4!6!} \cdot \frac{4!}{3!} \cdot \frac{1}{8} \right)$$

$$= 7!10! \cdot \left(\frac{1}{3!} + \frac{1}{6!4!} \cdot \frac{6}{2} + \frac{1}{5!} \cdot \frac{1}{3!2!} \cdot \frac{1}{4} + \frac{1}{6!3! \cdot 8} \right)$$

$$= \boxed{7,484,400}$$

(c) $10^7 = \boxed{10,000,000}$

(II) (a) $\binom{10}{7} = \boxed{120}$

(b) Clearly all outcomes that satisfy part (I)(a) also satisfy these conditions, so we start with $\binom{10}{7} = 120$ possible outcomes. In addition, the following outcomes are possible:

(i) A set of 6 children receive gifts; one child receives two gifts (6 distinct ways this could happen for each set of 6 children).

(ii) A set of 5 children receive gifts; two children receive two gifts ($\binom{5}{2}$ distinct ways this could happen for each set of 5 children).

(iii) A set of 4 children receive gifts; three children each receive two gifts (4 distinct ways this could happen for each set of 4 children).

Clearly each of these outcomes are mutually exclusive. Therefore the answer is

$$\binom{10}{7} + \binom{10}{6} \cdot \binom{6}{7-6} + \binom{10}{5} \cdot \binom{5}{7-5} + \binom{10}{4} \cdot \binom{4}{7-4} = \boxed{4,740}$$

(c) By Proposition 9, the number of nonnegative integer-valued vectors $(x_1, x_2, \ldots, x_r)$ satisfying the equation

$$x_1 + x_2 + \ldots + x_r = n$$

is equal to $\binom{n+r-1}{r-1}$. In distributing 7 identical gifts to 10 different children, we can imagine the vector $(x_1, x_2, \ldots, x_{10})$ represents the number of gifts given to each child (where x_i is a nonnegative integer for all i). So we have $n = 7$ and $r = 10$. Therefore the number of possible allocations is

$$\binom{7+10-1}{10-1} = \boxed{11,440}$$

Homework 1 Problem 2.

(I) 20 different gifts are distributed among seven children. How many different outcomes are possible if every child can receive (a) at least one gift, (b) at least two gifts, (c) any number of gifts?

(II) Answer the same questions if the gifts are identical (but the children are still different).

(III) Now try to generalize problems (1) and (2).

Solution.

(I) (a) There are 7^{20} possible allocations of gifts if we have no restrictions. If one child doesn't get a gift, there are $\binom{7}{1}$ ways to choose which child that is and 6^{20} subsequent allocations of gifts. Likewise, there are $\binom{7}{2} \cdot (7-2)^{20}$ ways to allocate the gifts if two children don't receive gifts, $\binom{7}{3} \cdot (7-3)^{20}$ ways if three children don't receive gifts, $\binom{7}{4} \cdot (7-4)^{20}$ ways if four children don't receive gifts, $\binom{7}{5} \cdot (7-5)^{20}$ ways if five children don't receive gifts, and $\binom{7}{6} \cdot (7-6)^{20}$ ways if six children don't receive gifts.

Let A_i denote the number of allocations in which i children do not receive gifts. In order to make the calculation, we must use the Inclusion-Exclusion principle (Proposition 5) (because, for example, some of the allocations in which three children don't receive gifts include allocations where four or more children don't receive gifts, and we don't want to double-count). Therefore the number of ways that at least one child can not receive a gift (i.e. the complement of every child receiving at least one gift) is

$$\left|\bigcup_{i=1}^{6} A_i\right| = \sum_{i=1}^{6} |A_i| - \sum_{1 \le i < j \le 6} |A_i \cap A_j| + \sum_{1 \le i < j < k \le 6} |A_i \cap A_j \cap A_k| - \ldots$$

$$+(-1)^{6-1} |A_1 \cap A_2 \cap A_3 \cap \cdots \cap A_6|$$

Fortunately, these allocations are nested in the sense that all the allocations where e.g. 5 children do not receive gifts are a subset of all the allocations where 4 children do not receive gifts; that is

$$A_6 \subset A_5 \subset A_4 \subset A_3 \subset A_2 \subset A_1$$

which implies e.g.

$$A_1 \cap A_2 \cap A_3 \cap \cdots \cap A_6 = A_6,$$

$$\sum_{1 \le i < j \le 6} |A_i \cap A_j| = 5|A_6| + 4|A_5| + 3|A_4| + 2|A_3| + |A_2|$$

So we have

$$\left|\bigcup_{i=1}^{6} A_i\right| = |A_6| + |A_5| + |A_4| + |A_3| + |A_2| + |A_1| - (5|A_6| + 4|A_5| + 3|A_4| + 2|A_3| + |A_2|)$$

$$+(4|A_6| + 3|A_5| + 2|A_4| + |A_3|) - (3|A_6| + 2|A_5| + |A_4|) + \ldots - |A_6|$$

$$= |A_1| - |A_2| + |A_3| - |A_4| + |A_5| - |A_6|$$

$$= \binom{7}{1} \cdot 6^{20} - \binom{7}{2} \cdot (7-2)^{20} + \binom{7}{3} \cdot (7-3)^{20} - \binom{7}{4} \cdot (7-4)^{20}$$

$$+\binom{7}{5} \cdot (7-5)^{20} - \binom{7}{6} \cdot (7-6)^{20}$$

The final answer is

$$7^{20} - \left|\bigcup_{i=1}^{6} A_i\right| = 7^{20} - \binom{7}{1} \cdot 6^{20} + \binom{7}{2} \cdot (7-2)^{20} - \binom{7}{3} \cdot (7-3)^{20} + \binom{7}{4} \cdot (7-4)^{20}$$

$$-\binom{7}{5} \cdot (7-5)^{20} + \binom{7}{6} \cdot (7-6)^{20} \approx \boxed{5.616 \cdot 10^{16}}$$

(b) Similar to above, but more complicated. The complement of every child receiving at least two gifts is that at least one child doesn't receive a gift (same as above) or at least one child only receives one gift. So we start from the baseline answer above, and subtract out all the possible allocations in which at least one child receives one gift.

If one child only receives one gift (and the rest receive more than one), there are $\binom{7}{1}$ ways to choose which child that is, $\binom{20}{1}$ ways to choose which gift that child receives, and 6^{20-1} allocations of the remaining gifts. If two children receive only one gift, there are $\binom{7}{2}$ ways to choose which children those are, $\binom{20}{2} \cdot 2!$ ways to choose which gifts those children get and distribute them among those children, and $(7-2)^{20-2}$ ways to allocate the remaining gifts. Likewise, if three children receive only one gift there are $\binom{7}{3}\binom{20}{3} \cdot 3! \cdot (7-3)^{20-3}$ ways to allocate the gifts, $\binom{7}{4}\binom{20}{4} \cdot 4! \cdot (7-4)^{20-4}$ ways if four children receive only one gift, $\binom{7}{5}\binom{20}{5} \cdot 5! \cdot (7-5)^{20-5}$ ways if five children receive only one gift, and $\binom{7}{6}\binom{20}{6} \cdot 6!(7-6)^{20-6}$ ways if six children don't receive gifts.

Let B_j be the event that j children receive only one gift. Note that $B_1 \cap A_i$ is nonempty $\forall i < 7-1$, $B_2 \cap A_i$ is nonempty $\forall i < 7-2$, and in general, $B_j \cap A_i$ is nonempty $\forall i < 7-j, j \in \{1, 2, \ldots, 6\}$. Applying the Inclusion-Exclusion Principle (Proposition 5) in a similar way as in part (I)(a), the answer is

$$\boxed{7^{20} - \left|\bigcup_{i=1}^{6} A_i\right| - \left|\bigcup_{j=1}^{6} B_j\right| + \sum_{i\in\{1,\ldots,6\}, j\in\{1,\ldots,6\}} \left|A_i \bigcap B_j\right|}$$

Per part (I)(a), the first two terms approximately equal $5.616 \cdot 10^16$. Clearly

$$\bigcup_{i\in\{1,\ldots,6\}, j\in\{1,\ldots,6\}} \left(A_i \bigcap B_j\right) \subset \bigcup_{j=1}^{6} B_j$$

which implies

$$-\left|\bigcup_{j=1}^{6} B_j\right| + \left|A_i \bigcap_{i\in\{1,\ldots,6\}, j\in\{1,\ldots,6\}} B_j\right| < 0$$

so the answer to this part will be less than $5.616 \cdot 10^16$, which makes sense.
Calculating $\left|\bigcup_{j=1}^{6} B_j\right|$ is not too difficult using Inclusion-Exclusion:

$$\left|\bigcup_{j=1}^{6} B_j\right| = \sum_{j=1}^{6} |B_j| - \sum_{1\leq j<k\leq 6} |B_j \cap B_k| + \sum_{1\leq j<k<\ell\leq 6} |B_j \cap B_k \cap B_\ell| - \ldots$$

$$+(-1)^{6-1} |B_1 \cap B_2 \cap B_3 \cap \cdots \cap B_6|$$

where since

$$B_6 \subset B_5 \subset B_4 \subset B_3 \subset B_2 \subset B_1$$

which implies e.g.

$$B_1 \cap B_2 \cap B_3 \cap \cdots \cap B_6 = B_6,$$

$$\sum_{1\leq j<k\leq 6} |B_j \cap B_j| = 5|B_6| + 4|B_5| + 3|B_4| + 2|B_3| + |B_2|$$

we have

$$\left|\bigcup_{j=1}^{6} B_j\right| = |B_6| + |B_5| + |B_4| + |B_3| + |B_2| + |B_1| - (5|B_6| + 4|B_5| + 3|B_4| + 2|B_3| + |B_2|)$$

$$+(4|B_6| + 3|B_5| + 2|B_4| + |B_3|) - (3|B_6| + 2|B_5| + |B_4|) + \ldots - |B_6|$$

$$= |B_1| - |B_2| + |B_3| - |B_4| + |B_5| - |B_6|$$

$$= \binom{7}{1}\binom{20}{1}\cdot(7-1)^{20-1} - \binom{7}{2}\binom{20}{2}\cdot 2!\cdot(7-2)^{20-2} + \binom{7}{3}\binom{20}{3}\cdot 3!\cdot(7-3)^{20-3} - \binom{7}{4}\binom{20}{4}\cdot 4!\cdot(7-4)^{20-4}$$

$$+\binom{7}{5}\binom{20}{5} \cdot 5! \cdot (7-5)^{20-5} - \binom{7}{6}\binom{20}{6} \cdot 6!(7-6)^{20-6}$$

$$\boxed{\approx 5.846 \cdot 10^{16}}$$

However, calculating

$$\sum_{i\in\{1,\ldots,6\}, j\in\{1,\ldots,6\}} \left|A_i \bigcap B_j\right|$$

is very difficult because, for example, $B_2 \cap A_3$ is nonempty but $B_2 \not\subset A_3$ and $A_3 \not\subset B_2$.

(c) $\boxed{7^{20} \approx 7.979 \cdot 10^{16}}$

(II) (a) Imagine we lay out the 20 indistinguishable gifts and take 6 sticks to place in the 19 spaces between them. The first child gets all the gifts to the left of the leftmost stick, the second child gets the gifts between the leftmost and second leftmost stick, and so on, until the last child gets all the gifts to the right of the rightmost stick. Because each child must receive at least one gift, we must place each stick in a different space. There are $\binom{19}{6} = \boxed{27,132}$ ways to do this.
(See also Proposition 8.)

(b) Similar to Problem 1 part (II)(c), if the vector $(x_1, x_2, \ldots, x_7)$ represents the number of gifts given to each child, we would like a solution such that

$$x_1 + x_2 + \ldots + x_7 = 20, x_i \geq 2 \ \forall i$$

Note that we can transform this problem in the following way:

$$x_1 - 1 + x_2 - 1 + \ldots + x_7 - 1 = 20 - 1 \cdot 7, x_i - 1 \geq 1 \ \forall i$$

Letting $y_i = x_i - 1$, we have the equivalent system

$$y_1 + y_2 + \ldots + y_7 = 13, y_i \geq 1 \ \forall i$$

By Proposition 8 (and because of the same logic as used in part (a)), the number of distinct solutions to this equation, and therefore the number of possible allocations under these conditions, is $\binom{12}{6} = \boxed{924}$.

(c) By Proposition 9, the number of nonnegative integer-valued vectors $(x_1, x_2, \ldots, x_r)$ satisfying the equation

$$x_1 + x_2 + \ldots + x_r = n$$

is equal to $\binom{n+r-1}{r-1}$. In distributing 20 identical gifts to 7 different children, we can imagine the vector $(x_1, x_2, \ldots, x_{10})$ represents the number of gifts given to each child (where x_i is a nonnegative integer for all i). So we have $n = 20$ and $r = 7$. Therefore the number of possible allocations is

$$\binom{20 + 7 - 1}{7 - 1} = \boxed{165,765,600}$$

(III) Generalization of 1(I): If there are g distinguishable gifts and $c \geq g$ children, the number of distinct allocations if each child can receive

(a) at most one gift is $\binom{c}{g} g!$.

(b) at most two gifts is

$$\sum_{i=c-g+1}^{g} \binom{c}{i} \cdot \binom{i}{g-i} \cdot \frac{g!}{(2!)^{g-i}}$$

(c) any number of gifts is c^g.

Generalization of 1(II): If there are g identical gifts and $c \geq g$ children, the number of distinct allocations if each child can receive

(a) at most one gift is $\binom{c}{g}$.

(b) at most two gifts is

$$\sum_{i=c-g+1}^{g} \binom{c}{i} \cdot \binom{i}{g-i}$$

(c) any number of gifts is $\binom{g+c-1}{c-1}$

Generalization of 2(I): If there are g distinguishable gifts and $c \leq g$ children, the number of distinct allocations if each child must receive

(a) at least one gift is

$$c^g - \sum_{i=1}^{c-1} (-1)^{i+1} \binom{c}{i} \cdot (c-i)^g$$

(b) at least two gifts is

$$c^g - \sum_{i=1}^{c-1} (-1)^{i+1} \binom{c}{i} \cdot (c-i)^g - \sum_{i=1}^{c-1} (-1)^{i+1} \binom{c}{i} \binom{g}{i} \cdot i! \cdot (c-i)^{g-i}$$

(c) any number of gifts is c^g

Generalization of 2(II): If there are g identical gifts and $c \leq g$ children, the number of distinct allocations if each child must receive

(a) at least one gift is

$$\binom{g-1}{c-1}$$

(b) at least two gifts is

$$\binom{g-c-1}{c-1}$$

(c) any number of gifts is

$$\binom{g+c-1}{c-1}$$

5.3 To Know for Math 505A Midterm 2

5.3.1 Definitions

Definition 5.15. A random variable X is **continuous** if its distribution function $F(x) = \Pr(X \leq x)$ can be written as

$$F(x) = \int_{-\infty}^{x} f(u) du$$

for some integrable $f : \mathbb{R} \to [0, \infty)$.

Definition 5.16. The function f is called the **(probability) density function** of the continuous random variable X.

Proposition 23. If X has pdf $f_X(x)$, then for $\mu \in \mathbb{R}$, $\sigma > 0$,

$$h(x) = \frac{1}{\sigma} f_X\left(\frac{x-\mu}{\sigma}\right)$$

is a pdf. In this setting μ is sometimes called a "location parameter" and σ is called a "scale parameter."

Definition 5.17. The **joint distribution function** of X and Y is the function $F : \mathbb{R}^2 \to [0, 1]$ given by

$$F(x, y) = \Pr(X \leq x \cap Y \leq y)$$

Definition 5.18. The random variables X and Y are **jointly continuous** with **joint (probability) density function** $f : \mathbb{R}^2 \to [0, \infty)$ if

$$F(x, y) = \int_{v=-\infty}^{y} \int_{u=-\infty}^{x} f(u, v) du dv \text{ for each } x, y \in \mathbb{R}$$

Definition 5.19. Two continuous random variables are **independent** if and only if $\{X \leq x\}$ and $\{Y \leq y\}$ are independent events for all $x, y \in \mathbb{R}$.

Ways to show independence:

- Use Definition 5.19: show that $\Pr(X \leq x \cap Y \leq y) = \Pr(X \leq x) \Pr(Y \leq y)$ for all $x, y \in \mathbb{R}$.
- **Theorem 24.** The random variables X and Y are independent if and only if $F(x, y) = F_X(x)F_Y(x)$ for all $x, y \in \mathbb{R}$.
- **Proposition 25.** For continuous random variables, the previous condition is equivalent to requiring $f(x, y) = f_X(x)f_Y(y)$.
- **Theorem 26.** If two variables are bivariate normal, they are independent if and only if their covariance

  $$\text{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xyf(x, y) dx dy$$

  is equal to 0.
- Characteristic functions:

  Theorem 27. X and Y are independent if and only if $\phi_{X,Y}(s, t) = \phi_X(s)\phi_Y(t)$.

Theorem 28. (Theorem 4.2.3, Grimmett and Stirzaker.) Let X and Y be random variables, and let $g, h : \mathbb{R} \to \mathbb{R}$. If X and Y are independent, then so are $g(X)$ and $h(Y)$.

5.3.2 Probability-Generating Functions

Definition 5.20.
$$G_X(s) = \mathbb{E}(s^X)$$

Theorem 29. Some useful properties:

(a) $\mathbb{E}(X) = G'_X(1)$, $\mathbb{E}[X(X-1)\cdots(X-k+1)] = G^{(k)}(1)$

(b) If X and Y are independent then $G_{X+Y}(s) = G_X(s)G_Y(s)$.

5.3.3 Moment-Generating Functions

Definition 5.21.
$$M_X(t) = \mathbb{E}(e^{tX})$$

Theorem 30. Some useful properties:

(a) $\mathbb{E}(X) = M'_X(0)$, $\mathbb{E}(X^k) = M^{(k)}(0)$

(b) If X and Y are independent then $M_{X+Y}(t) = M_X(t)M_Y(t)$.

5.3.4 Characteristic Functions

Definition 5.22.
$$\phi_X(t) = \mathbb{E}(e^{itX})$$

Proposition 31. Necessary and sufficient conditions for a function to be a characteristic function:

(a) $\phi_X(0) = 1$

(b) $|\phi(t)| \leq 1 \;\forall\; t$

(c) ϕ is uniformly continuous on $\mathbb{R}$

(d) ϕ is positive semidefinite; that is,

$$\sum_{i,j} \phi(t_j - t_k) z_j \overline{z}_k \geq 0 \text{ for all real } t_1, t_2, \ldots, t_n \text{ and complex } z_1, z_2, \ldots, z_n$$

Or, equivalently, or every set of real numbers $t_1, t_2, \ldots, t_n$, the matrix $\phi(t_i - t_j), i, j \in \{1, 2, \ldots, n\}$ is Hermitian and nonnegative definite.

Remark. Relationship between characteristic functions and probability and moment generating functions:

$$\phi_X(t) = M_X(it) = G_X(e^{it})$$

Theorem 32. Some useful properties:

(a) $X \perp\!\!\!\perp Y \implies \phi_{X+Y}(t) = \phi_X(t)\phi_Y(t)$

(b) $Y = aX + b \implies \phi_Y(t) = e^{itb}\phi_X(at)$

(c) $\phi_X^{(k)}(0) = i^k \mathbb{E}(X^k)$

(d) $\phi_{X,Y}(s,t) = \mathbb{E}(e^{isX}e^{itY})$

(e) $X \perp\!\!\!\perp Y \iff \phi_{X,Y}(s,t) = \phi_X(s)\phi_Y(t)$

Theorem 33. Other facts from notes on course website

(a) If $\phi(t)$ is even, $\phi(0) = 1$, ϕ is convex for $t > 0$, and $\lim_{t\to\infty}\phi(t) = 0$, then ϕ is a characteristic function of an absolutely continuous random variable.

(b) If ϕ is a characteristic function and $\phi(t) = 1 + o(t^2), t \to 0$, then $\phi(t) = 1$ for all t. The random variable with such a characteristic function must have zero mean and zero variance. In particular, if $r > 2$, then $\exp(-|t|^r)$ is not a characteristic function.

(c) If $\phi(t) = e^{p(t)}$ is a characteristic function and $p = p(t)$ is a polynomial, then the degree of p is at most 2. For example, $e^{t^2-t^4}$ is not a characteristic function.

(d) If ξ is absolutely continuous, then $\lim_{|t|\to\infty}|\phi_\xi(t)| = 0$ (Riemann-Lebesgue).

(e) If $\int_{-\infty}^{\infty}|\phi_\xi(t)|\,dt < \infty$, then ξ is absolutely continuous with pdf

$$f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\exp(-itx)\phi(t)dt$$

5.3.5 Continuous Random Variable Distributions

Uniform: $\mathrm{U}(a,b)$

- Probability density function:

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

- Cumulative distribution function:

$$F(x) = \Pr(X \le x) = \begin{cases} 0 & x \le a \\ \frac{x-a}{b-a} & a < x \le b \\ 1 & x > b \end{cases}$$

- Probability-generating function:
- Moment-generating function:
- Characteristic function:

- Expectation: $\mathbb{E}(X) = (b-a)/2$
- Variance: $\text{Var}(X) = (b-a)^2/12$

Normal: $\mathcal{N}(\mu, \sigma^2)$

- Probability density function:

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(- \frac{(x-\mu)^2}{2\sigma^2} \right)$$

- Cumulative distribution function: $F(x) = \Pr(X \leq x) =$
- Probability-generating function:
- Moment-generating function:
- Characteristic function: $\phi(t) = \exp(i\mu t - (1/2)\sigma^2 t^2)$. Standard normal: $\phi(t) = \exp((-1/2)t^2)$.
- Expectation: $\mathbb{E}(X) = \mu$
- Variance: $\text{Var}(X) = \sigma^2$

Gamma: $\Gamma(\alpha, \beta)$ (**note: this parameterization is a little unusual; more commonly β is expressed as the reciprocal of how it appears here.**)

- Probability density function:

$$f(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta} = \frac{1}{\Gamma(\alpha, \beta)} x^{\alpha-1} e^{-x/\beta}$$

- Cumulative distribution function: $F(x) = \Pr(X \leq x) =$
- Probability-generating function:
- Moment-generating function:
- Characteristic function:
- Expectation: $\mathbb{E}(X) = \alpha\beta$
- Variance: $\text{Var}(X) = \alpha\beta^2$

Proposition 34. Let $X \sim \text{Gamma}(\alpha_1, \beta)$ and $Y \sim \text{Gamma}(\alpha_2, \beta)$. Then $X + Y \sim \text{Gamma}(\alpha_1 + \alpha_2, \beta)$.

χ^2_n: special case of a gamma distribution: $\Gamma(n/2, 2)$. Also the sum of n independent standard normally distributed variables.

- Probability density function:

$$f(x) = \frac{1}{2^{n/2}\Gamma(n/2)} x^{n/2-1} e^{-x/2} = \frac{1}{\Gamma(n/2, 2)} x^{n/2-1} e^{-x/2}$$

- Cumulative distribution function: $F(x) = \Pr(X \leq x) =$
- Probability-generating function:
- Moment-generating function:
- Characteristic function:
- Expectation: $\mathbb{E}(X) = n/2 \cdot 2 = n$
- Variance: $\text{Var}(X) = n/2 \cdot 2^2 = 2n$

Exponential: (special case of a gamma distribution: $\Gamma(1, \beta)$. Also a special case of a Weibull distribution with $\beta = 1$.)

- Probability density function: $f(x) = \frac{1}{\beta} \exp(-x/\beta) = \lambda e^{-\lambda x}$
- Cumulative distribution function: $F(x) = \Pr(X \leq x) = 1 - e^{-\lambda x}$
- Probability-generating function:
- Moment-generating function:
- Characteristic function:
- Expectation: $\mathbb{E}(X) = \beta = \lambda^{-1}$
- Variance: $\text{Var}(X) = \beta^2 = \lambda^{-2}$

Cauchy:

- Probability density function:

$$f(x) = \frac{1}{\pi(1 + x^2)} \text{ (standard Cauchy) }, f(x) = \frac{1}{\pi\sigma(1 + (x - \mu)^2/\sigma^2)} \text{ (general)}$$

- Cumulative distribution function: $F(x) = \Pr(X \leq x) =$
- Probability-generating function:
- Moment-generating function:
- Characteristic function:
- Expectation: does not exist
- Variance: does not exist (Cauchy distribution has no moments.)

Beta: Recall:

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)}$$

$$\implies \frac{B(\alpha + 1, \beta)}{B(\alpha, \beta)} = \frac{\Gamma(\alpha + 1)\Gamma(\beta)}{\Gamma(\alpha + 1 + \beta)} \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} = \frac{\alpha}{\alpha + \beta}$$

- Probability density function: $f(x) =$
- Cumulative distribution function: $F(x) = \Pr(X \leq x) =$
- Probability-generating function:
- Moment-generating function:
- Characteristic function:
- Expectation: $\mathbb{E}(X) =$
- Variance: $\text{Var}(X) =$

t_n:

- Probability density function:

$$f(x) = \frac{\Gamma((n+1)/2)}{\sqrt{n\pi} \cdot \Gamma(n/2)} \left(1 + \frac{x^2}{n}\right)^{-(n+1)/2}$$

- Cumulative distribution function: $F(x) = \Pr(X \leq x) =$
- Probability-generating function:
- Moment-generating function:
- Characteristic function:
- Expectation: $\mathbb{E}(X) = 0$
- Variance: $\text{Var}(X) = n/(n-2)$

Weibull:

- Probability density function: $f(x) = \alpha\beta x^{\beta-1} \exp(-\alpha x^\beta)$
- Cumulative distribution function: $F(x) = \Pr(X \leq x) = 1 - \exp(-\alpha x^\beta)$
- Probability-generating function:
- Moment-generating function:
- Characteristic function:
- Expectation: $\mathbb{E}(X) =$
- Variance: $\text{Var}(X) =$

5.3.6 Multivariate Gaussian (Normal) Distributions

Definition 5.23. From http://pluto.huji.ac.il/~pchiga/teaching/MathStat/SIAnotes2013.pdf (definition 2b6): A random vector $X = (X_1, X_2)$ is Gaussian with mean $\mu = (\mu_1, \mu_2)$ and the covariance matrix

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$$

if it has a joint pdf of the form

$$f_X(x) = \frac{1}{2\pi\sigma_2\sigma_2\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2}\frac{1}{1-\rho^2}\left(\frac{(x_1-\mu_1)^2}{\sigma_1^2)} - \frac{2\rho(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_1} + \frac{(x_2-\mu_2)^2}{\sigma_2^2)} \right)\right]$$

for $x \in \mathbb{R}^2$.

Proposition 35. From http://pluto.huji.ac.il/~pchiga/teaching/MathStat/SIAnotes2013.pdf (Proposition 3c1): Let X be a Gaussian random variable in $\mathbb{R}^2$ as in Definition 2b6. Then $f_{X_1|X_2}(x_1; x_2)$ is Gaussian with the (conditional) mean

$$\mathbb{E}(X_1 \mid X_2 = x_2) = \mu_1 + \frac{\rho\sigma_1}{\sigma_2}(x_2 - \mu_2)$$

and the (conditional) variance

$$\mathrm{Var}(X_1 \mid X_2 = x_2) = \sigma_1^2(1-\rho^2)$$

Recall Theorem 26: if two variables are bivariate normal, they are independent if and only if their covariance

$$\mathrm{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf(x,y)dxdy$$

equals 0.

Theorem 36. For a bivariate normal distribution

$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix} \right)$$

the conditional distribution of X_1 given X_2 is

$$X_1 \mid X_2 = x_2 \sim \mathcal{N}\left(\mu_1 + \rho\frac{\sigma_1}{\sigma_2}(x_2 - \mu_2), (1-\rho^2)\sigma_1^2 \right)$$

Remark. Note that this matches the OLS coefficients in the univariate case. In other words, the univariate OLS formula can be derived using only this fact.

5.4 Worked problems

5.4.1 Example Problems That Will Likely Appear on Midterm

(1) **Note: we worked through an example problem like this on Friday. Should probably fix solution, and use geometric**

Question:

Let X, Y, Z be independent uniform on (0, 1). Compute the cdfs of XY, X/Y, and XY/Z.

Solution (may not be the way Lototsky suggested, consider revising).

Using the information from part (a), and the fact that $f_X(x) = 1$ (for $x \in [0,1]$) and likewise for $f_Y(y)$:

- XY:

$$F_{XY}(z) = \int_0^\infty f_X(x) \int_{-\infty}^{z/x} f_Y(y) dy dx - \int_{-\infty}^0 f_X(x) \int_\infty^{z/x} f_Y(y) dy dx$$

$$= \int_0^1 \left[(z/x)\mathbf{1}_{\{0<z/x\le 1\}} + \mathbf{1}_{\{z/x>1\}}\right] dx = \int_0^1 \left[(z/x)\mathbf{1}_{\{z\le x\}} + \mathbf{1}_{\{z>x\}}\right] dx = \int_0^z dx + \int_z^1 (z/x) dx$$

$$= z + z\log(x)\Big|_z^1 = z + z\log(1) - z\log(z) = z(1-\log(z))$$

$$\implies \boxed{F_{XY}(z) = \begin{cases} 0 & z \le 0 \\ z(1-\log(z)) & 0 < z \le 1 \\ 1 & z > 1 \end{cases}}$$

- X/Y:

$$F_{X/Y}(z) = \int_0^\infty f_Y(y) \int_{-\infty}^{zy} f_X(x) dx dy - \int_{-\infty}^0 f_Y(y) \int_\infty^{zy} f_X(x) dx dy$$

$$= \int_0^1 \left[zy\mathbf{1}_{\{0<zy\le 1\}} + \mathbf{1}_{\{zy>1\}}\right] dy = \int_0^1 \left[zy\mathbf{1}_{\{y>0 \cap y\le 1/z\}} + \mathbf{1}_{\{y>1/z\}}\right] dy = \int_0^{1/z} zy \cdot dy + \int_{1/z}^1 dy$$

$$= \frac{zy^2}{2}\Big|_0^{1/z} + (1 - 1/z) = \frac{z}{2z^2} + 1 - \frac{2}{2z} = 1 - \frac{1}{2z}$$

$$\implies F_{XY}(z) = \begin{cases} 0 & z \le 0 \\ 1 - \frac{1}{2z} & 0 < 1/z \le 1 \end{cases}$$

$$= \boxed{\begin{cases} 0 & z \le 0 \\ z/2 & 0 < z \le 1 \\ 1 - \frac{1}{2z} & z > 1 \end{cases}}$$

- XY/Z: Consider this the cdf of the quotient of $W = XY$ and Z.

$$F_U(u) = \int_0^\infty f_Z(z) \int_{-\infty}^{uz} f_W(w) dw dz - \int_{-\infty}^0 f_Z(z) \int_\infty^{uz} f_W(w) dw dz$$

$$= \int_0^1 \int_0^{uz} -\log(w) \mathbf{1}_{\{0<uz\leq 1\}} dw dz = \int_0^1 -\big[w\log(w) - w\big]_0^{uz} \mathbf{1}_{\{0<z\leq 1/u\}} dz$$

$$= \int_0^{1/u} uz\big[1 - \log(uz)\big] dz = \frac{u}{4} z^2 \big(3 - 2\log(uz)\big)\Big|_0^{1/u} = \frac{u}{4u^2}\big(3 - 2\log(1)\big) - 0 = \frac{3}{4u}$$

$$\implies \boxed{F_{XY/Z}(u) = \begin{cases} 0 & u \leq 0 \\ \frac{3}{4u} & 0 < u \leq 3/4 \\ 1 & u > 3/4 \end{cases}}$$

(2) **Note: we worked through an example problem like this on Friday.**

Question from Friday: Let X, Y be distributed exponentially with mean 1. What is the probability distribution of $X/(X+Y)$?

Solution. Find the cdf:

$$\Pr\left(\frac{X}{X+Y} < t\right) = \Pr(X < tX + tY) = \Pr\left(Y > \frac{X(1-t)}{t}\right)$$

To find $\Pr(Y > aX)$, graph the line aX

5.4.2 More Problems From Homework

Homework 5 Problem 4.

Let $X_1, X_2, \ldots$ be i.i.d. having moment-generating functions $M_X = M_X(t), t \in (-\infty, \infty)$. Let N be an integer-valued random variable with moment-generating function $M_N = M_N(t), t \in (-\infty, \infty)$. Assume that N is independent of all X_k and define $S = \sum_{k=1}^N X_k$. Confirm that the random variable S has the moment-generating function $M_S = M_S(t)$ defined for all $t \in (-\infty, \infty)$ and

$$M_S(t) = M_N\big(M_X(t)\big)$$

Then use the result to derive the formulae

$$\mathbb{E}(S) = \mu_N \mu_X, \mathrm{Var}(S) = (\sigma_N^2 - \mu_N)\mu_X^2 + \mu_N \sigma_X^2$$

where $\mu_N = \mathbb{E}(N)$, $\mu_X = \mathbb{E}(X_1)$, $\sigma_N^2 = \mathrm{Var}(N)$, and $\sigma_X^2 = \mathrm{Var}(X_1)$. How will the above computations change if we use the characteristic function ϕ_X instead of the moment-generating function M_X?

Solution.

$$M_S(t) = \mathbb{E}(e^{tS}) = \mathbb{E}[\mathbb{E}(e^{tS} \mid N)] = \sum_{n=0}^{\infty} \mathbb{E}(e^{tS} \mid N = n) \Pr(N = n) = \sum_{n=0}^{\infty} \mathbb{E}(e^{t(X_1+X_2+\ldots+X_n)} \mid N = n) \Pr(N = n)$$

$$= \sum_{n=0}^{\infty} \mathbb{E}(e^{tX_1} e^{tX_2} \cdots e^{tX_n}) \Pr(N = n)$$

By independence of the X_i we have

$$= \sum_{n=0}^{\infty} \mathbb{E}(e^{tX_1})\mathbb{E}(e^{tX_2}) \cdots \mathbb{E}(e^{tX_n}) \Pr(N = n)$$

which, since the X_i are i.i.d., can be written as

$$= \sum_{n=0}^{\infty} \mathbb{E}(e^{tX_1})^n \Pr(N = n) = \sum_{n=0}^{\infty} (M_X(t))^n \Pr(N = n)$$

But since $G_N(s) = \mathbb{E}(s^N) = \sum_{n=0}^{\infty} s^n \Pr(N = n)$, this can be written as

$$M_S(t) = G_N(M_X(t))$$

as desired. Note that

$$M_S'(t) = G_N'\big(M_X(t)\big)M_X'(t)$$

$$M_S''(t) = G_N''\big(M_X(t)\big)(M_X'(t))^2 + G_N'\big(M_X(t)\big)M_X''(t)$$

So we have

- $\mathbb{E}(S) = M_S'(0) = G_N'\big(M_X(0)\big)M_X'(0) = G_N'(1)\mathbb{E}(X_1) = \mathbb{E}(N)\mathbb{E}(X_1) = \mu_N \mu_X$
- $\mathrm{Var}(S) = \mathbb{E}(S^2) - \mathbb{E}(S)^2 = M_S''(0) - (M_S'(0))^2$

$$= G_N''\big(M_X(0)\big)(M_X'(0))^2 + G_N'\big(M_X(0)\big)M_X''(0) - \mu_N^2\mu_X^2 = G_N''\big(1\big)\mathbb{E}(X_1)^2 + G_N'\big(1\big)\mathrm{Var}(X_1) - \mu_N^2\mu_X^2$$

$$= \mathbb{E}[N(N-1)]\mathbb{E}(X_1)^2 + \mathbb{E}(N)\mathrm{Var}(X_1) - \mu_N^2\mu_X^2 = \mathbb{E}[N^2 - N]\mathbb{E}(X_1)^2 + \mathbb{E}(N)\mathrm{Var}(X_1) - \mu_N^2\mu_X^2$$

$$= [\mathbb{E}(N^2) - \mathbb{E}(N)^2 + \mathbb{E}(N)^2 - \mathbb{E}(N)]\mathbb{E}(X_1)^2 + \mathbb{E}(N)\mathrm{Var}(X_1) - \mu_N^2\mu_X^2 =$$

$$= [\text{Var}(N) + \mathbb{E}(N)^2 - \mathbb{E}(N)]\mathbb{E}(X_1)^2 + \mathbb{E}(N)\text{Var}(X_1) - \mu_N^2\mu_X^2 = (\sigma_N^2 + \mu_N^2 - \mu_N)\mu_X^2 + \mu_N\sigma_X^2 - \mu_N^2\mu_X^2$$

$$= \boxed{(\sigma_N^2 - \mu_N)\mu_X^2 + \mu_N\sigma_X^2}$$

To use the characteristic function ϕ_X instead of the moment generating function M_X, we would do the following:

$$\phi_S(t) = \mathbb{E}(e^{itS}) = \mathbb{E}[\mathbb{E}(e^{itS} \mid N)] = \sum_{n=0}^{\infty} \mathbb{E}(e^{itS} \mid N = n)\Pr(N = n) = \sum_{n=0}^{\infty} \mathbb{E}(e^{it(X_1+X_2+\ldots+X_n)} \mid N = n)\Pr(N = n)$$

$$= \sum_{n=0}^{\infty} \mathbb{E}(e^{itX_1}e^{tX_2}\cdots e^{itX_n})\Pr(N = n)$$

By independence of the X_i we have

$$= \sum_{n=0}^{\infty} \mathbb{E}(e^{itX_1})\mathbb{E}(e^{itX_2})\cdots\mathbb{E}(e^{itX_n})\Pr(N = n)$$

which, since the X_i are i.i.d., can be written as

$$= \sum_{n=0}^{\infty} \mathbb{E}(e^{itX_1})^n\Pr(N = n) = \sum_{n=0}^{\infty} (\phi_X(t))^n\Pr(N = n)$$

But since $G_N(s) = \mathbb{E}(s^N) = \sum_{n=0}^{\infty} s^n\Pr(N = n)$, this can be written as

$$\phi_S(t) = G_N(\phi_X(t))$$

Homework 5 Problem 7.

(a) Let $X_1, X_2, \ldots, X_n$ be independent with mean zero and finite third moment. Prove that

$$\mathbb{E}(X_1 + \ldots + X_n)^3 = \mathbb{E}X_1^3 + \ldots + \mathbb{E}X_n^3$$

Solution.

(a) Let $\mathbb{E}(\exp(it_iX_i) = \phi_{X_i}(t_i)$. Let $S_n = \sum_{i=1}^n X_i$. Then by independence the characteristic function for S_n is

$$\mathbb{E}(\exp(itS_n)) = \phi_{S_n}(t) = \prod_{i=1}^n \phi_{X_i}(t)$$

Then

$$\mathbb{E}(X_1 + X_2 + \ldots + X_n)^3 = \mathbb{E}(S_n^3) = \phi_{S_n}^{(3)}(0)$$

$$= \sum_{i=1}^n \phi_{X_i}^{(3)}(0) \cdot \left(\prod_{j \in \{1,\ldots,n\}, j \neq i} \phi_{X_j}(0) \right) + C \left[\sum_{i=1}^n \cdot \left(\sum_{j \in \{1,\ldots,n\}, j \neq i} \phi_{X_i}^{(2)}(0) \phi_{X_j}^{(1)}(0) \right) \cdot \left(\prod_{k \in \{1,\ldots,n\}, k \neq i,j} \phi_{X_k}(0) \right) \right]$$

where C is some coefficient resulting from the multinomial expansion of S_n after repeated differentiation product rules. But because $\mathbb{E}(X_i) = 0$, $\phi_{X_i}^{(1)}(0) = 0 \ \forall i$, so the second term goes to 0. Therefore we have

$$\mathbb{E}(X_1 + X_2 + \ldots + X_n)^3 = \sum_{i=1}^n \phi_{X_i}^{(3)}(0) \cdot \left(\prod_{j \in \{1,\ldots,n\}, j \neq i} \phi_{X_j}(0) \right) = \sum_{i=1}^n \mathbb{E}(X_i^3) \cdot 1^{n-1} = \sum_{i=1}^n \mathbb{E}(X_i^3)$$

as desired.

Homework 6 Problem 10.

(a) For $p \in (0,1)$, let $x(p)$ be the smallest number of people so that there is a better than $100 \cdot p\%$ chance to have at least two born on the same day. Find an approximate expression for $x(p)$, and sketch the graph of the function $x = x(p)$.

(b) Repeat part (a) when you want at least three people to share a birthday.

Solution.

(a) Let $f(x)$ be the probability of no matches in birthdays in a group of x people; that is,

$$f(x) = \frac{365 \cdot 364 \cdot 363 \cdots (365 - x + 1)}{365^x} = \frac{1}{365^x} \cdot \frac{365!}{(365-x)!} = \left(1 - \frac{1}{365}\right)\left(1 - \frac{2}{365}\right) \cdots \left(1 - \frac{x-1}{365}\right)$$

Using the first order Taylor approximation $\exp(-k/x) \approx 1 - k/x$, we have

$$f(x) = \left(1 - \frac{1}{365}\right)\left(1 - \frac{2}{365}\right) \cdots \left(1 - \frac{x-1}{365}\right) \approx \exp(-1/365)\exp(-2/365) \cdots \exp(-(x-1)/365)$$

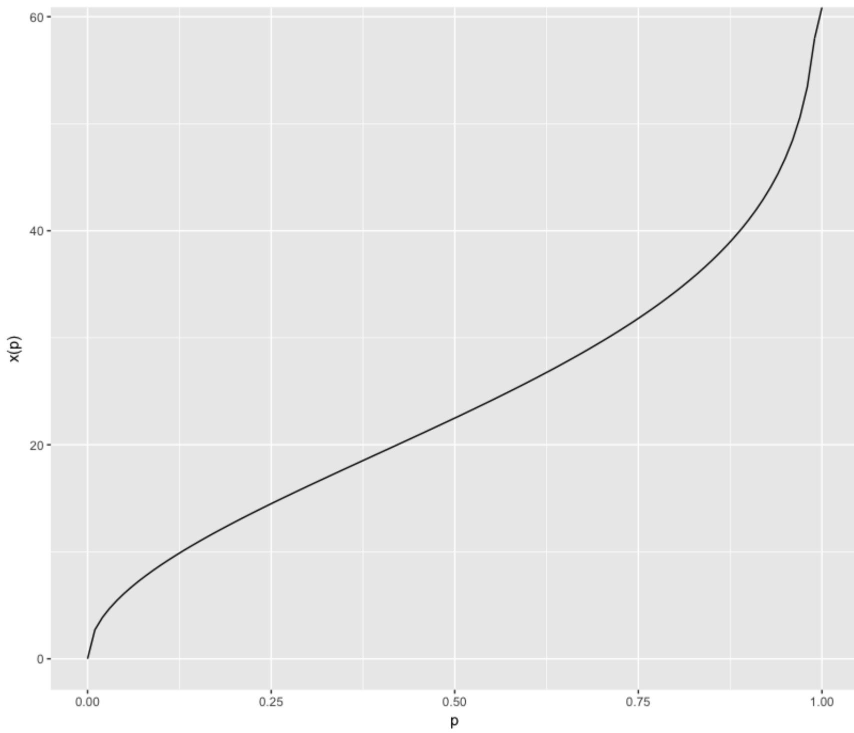
$$= e^{-(x^2 - x)/(2 \cdot 365)}$$

We want the probability of a match to be at least p; that is, $f(x) \leq 1 - p$. Setting this equal to $q = 1 - p$, we have

$$e^{-(x^2-x)/(2\cdot 365)} = q \iff -\frac{x^2 - x}{730} = \log(q) \iff x^2 - x + 730 \log(q) = 0$$

$$\implies x = 0.5 + \sqrt{1/4 + 730\log(1/q)} \approx \boxed{\sqrt{2 \cdot 365\log(1/q)}}$$

where we discard the negative root because we have to have a nonnegative number of people, and we don't worry about the decimals since this is an approximation and we have to round up to the nearest whole person anyway.



(b) For a group of three people, the Poisson approximation is more convenient. The number of groups of 3 people in a room of x people is $\binom{x}{3}$. For a group of three people, the probability that all three have the same birthday is $1 \cdot 1/365 \cdot 1/365 = 365^{-2}$. Therefore we can think of the number of matches of three people as distributed Poisson with expectation $\binom{x}{3} \cdot 365^{-2}$. Then we have the probability of at least one "success" (triplet with three matched birthdays) is

$$1 - \frac{\exp(-\lambda)\lambda^0}{0!} = 1 - \exp\left(-\binom{x}{3} \cdot 365^{-2}\right)$$

We set this equal to p and solve:

$$p = 1 - \exp\left(-\binom{x}{3} \cdot 365^{-2}\right) \iff -\binom{x}{3} \cdot 365^{-2} = \log(1-p) \iff \frac{x!}{(x-3)!3!} = 365^2 \cdot \log\left(\frac{1}{1-p}\right)$$

$$\iff x(x-1)(x-2) = 6 \cdot 365^2 \cdot \log\left(\frac{1}{1-p}\right) \iff (x^2-x)(x-2) = x^3 - 3x^2 + 2x = 6 \cdot 365^2 \cdot \log\left(\frac{1}{1-p}\right)$$

This has a unique real solution, but it is hard to find.

6 Stochastic Processes

These notes are based on my notes from *Time Series and Panel Data Econometrics* (1st edition) by M. Hashem Pesaran as well as coursework for Economics 613: Economic and Financial Time Series I at USC.

6.1 Martingales

Definition. Let $\{y_t\}_{t=0}^{\infty}$ be a sequence of random variables, and let Ω_t denote the information set available at date t, which at least contains $\{y_t, y_{t-1}, y_{t-2}, \ldots\}$. If $\mathbb{E}(y_t \mid \Omega_{t-1}) = y_{t-1}$ holds then $\{y_t\}$ is a martingale process with respect to Ω_t.

Definition. Let $\{y_t\}_{t=1}^{\infty}$ be a sequence of random variables, and let Ω_t denote the information set available at date t, which at least contains $\{y_t, y_{t-1}, y_{t-2}, \ldots\}$. If $\mathbb{E}(y_t \mid \Omega_{t-1}) = 0$, then $\{y_t\}$ is a martingale difference process with respect to Ω_t.

6.2 Brownian Motion

Appendix B.13, Brownian motion. A standard Brownian motion $b(\cdot)$ is a continuous-time stochastic process associating each date $a \in [0,1]$ with the scalar $b(a)$ such that

(i) b(0) = 0

(ii) For any dates $0 \leq a_1 \leq a_2 \leq \ldots \leq a_k \leq 1$ the changes $[b(a_2) - b(a_1)]$, $[b(a_3) - b(a_2)], \ldots, [b(a_k) - b(a_k - 1)]$ are independent multivariate Gaussian with $b(a) - b(s) \sim \mathcal{N}(0, a - s)$.

(iii) For any given realization, $b(a)$ is continuous in a with probability 1.

Other continuous time processes can be generated from the standard Brownian motion. For example, a Brownian motion with variance σ^2 can be obtained as

$$w(a) = \sigma b(a)$$

where $b(a)$ is a standard Brownian motion.

The continuous time process

$$\boldsymbol{w}(a) = \boldsymbol{\Sigma}^{1/2} \boldsymbol{b}(a)$$

is a Brownian motion with covariance matrix $\boldsymbol{\Sigma}$.

Definition 26 (Wiener process). Let $\Delta w(t)$ be the change in $w(t)$ during the time interval dt. Then $w(t)$ is said to follow a Wiener process if

$$\Delta w(t) = \epsilon_t \sqrt{dt}, \quad \epsilon_t \sim IID(0,1)$$

and $w(t)$ denotes the value of the $w(\cdot)$ at date t. Clearly,

$$\mathbb{E}[\Delta w(t)] = 0, \text{ and } \operatorname{Var}[\Delta w(t)] = dt$$

Donsker's Theorem, Theorem 43, p.335, Section 15.6.3. Let $a \in [0,1)$, $t \in [0,T]$, and suppose $(J-1)/T \leq a < J/T, J = 1, 2, \ldots, T$. Define

$$R_T(a) = \frac{1}{\sqrt{T}} s_{\left[Ta\right]}$$

where

$$s_{\left[Ta\right]} = \epsilon_1 + \epsilon_2 + \ldots + \epsilon_{\left[Ta\right]}$$

$\left[Ta\right]$ denotes the largest integer part of Ta and $s_{\left[Ta\right]} = 0$ if $\left[Ta\right] = 0$. Then $R_T(a)$ weakly converges to $w(a)$, i.e.,

$$R_T(a) \to w(a)$$

where $w(a)$ is a Wiener process. Note that when $a = 1$, $R_T(1) = 1/\sqrt{T} \cdot S_{\left[T\right]} = 1/\sqrt{T} \cdot (\epsilon_1 + \epsilon_2 + \ldots + \epsilon_T$. Since ϵ_t's are IID, by the central limit theorem, $R_T(1) \to \mathcal{N}(0,1)$.

Similar (Theorem 2.1 in Phillips and Durlaf (1986)): Let $\{u_t\}$ be a sequence satisfying $\mathbb{E}(u_t) = 0$, $\gamma(0) = \mathbb{E}(T^{-1}S_t^2) \to \sigma^2 < \infty$ as $T \to \infty$, $\{u_t\}$ is square summable, $\sup_t\{\mathbb{E}(|u_t|^\beta)\} < \infty$ for some $2 \leq \beta < \infty$ and all t, $\gamma(h) = \mathbb{E}(T^{-1}(y_t - y_{t-h})^2) \to K_h < \infty$ as $\min\{h, T\} \to \infty$. Then $X_T(t) \implies W(t)$ as $T \to \infty$, where $W(t)$ is a Wiener process.

Theorem 37. Continuous Mapping Theorem (Theorem 44 of Pesaran in 15.6.3). Let $a \in [0,1)$, $i \in [0,n]$, and suppose $(J-1)/n \leq a < J/n, J = 1, 2, \ldots, n$. Define $R_n(a) = n^{-1/2} S_{\left[n \cdot a\right]}$. If $f(\cdot)$ is continuous over $[0,1)$, then

$$f[R_n(a)] \xrightarrow{d} f[w(a)]$$

7 Asymptotics and Convergence

These notes are based on my notes from chapter 8 of *Time Series and Panel Data Econometrics* (1st edition) by M. Hashem Pesaran and coursework for Economics 613: Economic and Financial Time Series I at USC, as well as Math 505A at USC and chapter 7 from *Probability and Random Processes* (Grimmet and Stirkazer) 3rd edition.

7.1 Preliminaries (5.9 and 7.1, Grimmett and Stirzaker)

Definition 7.1. Definition 7.1.4, Grimmett and Stirzaker. If for all $x \in [0, 1]$ the sequence $\{f_n(x)\}$ of real numbers satisfies $f_n(x) \to f(x)$ as $n \to \infty$ then we say $f_n \to f$ **pointwise.**

Remark. In practice pointwise convergence is often not useful for functions because a sequence of functions may be continuous while its limit is not. For instance, consider $\{f_n : f_n = x^n \ \forall x \in [0, 1]\}$. Then f_n is continuous for all n but

$$\lim_{n\to\infty} f_n = \begin{cases} 0 & x \leq 1 \\ 1 & x = 1 \end{cases}$$

Instead, the following definition is often more useful.

Definition 7.2. (from class notes.) We say that f_n **uniformly converges to** f **on** $[a, b]$ if for every $\epsilon > 0$ there exists N such that for every $n > N$,

$$\forall x \in [a, b] \ |f_n(x) - f(x)| < \epsilon$$

Definition 7.3. (Definition 7.1.5, Grimmett and Stirzaker.) Let V be a collection of functions mapping $[0, 1]$ into $\mathbb{R}$ and assume V is endowed with a function $\|\cdot\| : V \to \mathbb{R}$ satisfying

(a) $\|f\| \geq 0$ for all $f \in V$

(b) $\|f\| = 0$ if and only if f is the zero function (or equivalent to it)

(c) $\|af\| = |a| \cdot \|f\|$ for all $a \in \mathbb{R}$, $f \in V$

(d) $\|f + g\| \leq \|f\| + \|g\|$ (triangle inequality)

The function $\|\cdot\|$ is called a **norm**. If $\{f_n\}$ is a sequence of members of V then we say that $f_n \to f$ **with respect to the norm** $\|\cdot\|$ if $\|f_n - f\| \to 0$ as $n \to \infty$.

Definition 7.4. (Definition 7.16, Grimmett and Stirzaker.) Let $\epsilon > 0$ be prescribed, and define the distance between two functions $g, h : [0, 1] \to \mathbb{R}$ by

$$d_\epsilon(g, h) = \int_E dx$$

where $E = \{u \in [0, 1] : |g(u) - h(u)| > \epsilon\}$. We say that $f_n \to f$ **in measure** if

$$d_\epsilon(f_n, f) \to 0 \text{ as } n \to \infty \text{ for all } \epsilon > 0$$

Theorem 38. Inversion Theorem (Theorem 5.9.2, Grimmett and Stirzaker). Let X have distribution function F and characteristic function ϕ. Define $\overline{F} : \mathbb{R} \to [0, 1]$ by

$$\overline{F}(x) = \frac{1}{2}\left[F(x) + \lim_{y \to x^-} F(y)\right]$$

Then

$$\overline{F}(b) - \overline{F}(a) = \lim_{N \to \infty} \int_{-N}^{N} \frac{\exp(-iat) - \exp(-ibt)}{2\pi it} \cdot \phi(t) dt$$

Proof. See Kingman and Taylor (1966). □

Corollary 38.1. Corollary 5.9.3. Random variables X and Y have the same characteristic function if and only if they have the same distribution function.

Proof. **Available in Grimmett and Stirzaker section 5.9, pp. 189 - 190.** □

Definition 7.5. (Definition 5.9.4, Grimmett and Stirzaker.) We say that the sequence $F_1, F_2, \ldots$ of distribution functions **converges** to the distribution function F (written $F_n \to F$) if $F(x) = \lim_{n\to\infty} F_n(x)$ at each point x where F is continuous.

Theorem 39. Continuity theorem (Thereom 5.9.5; in notes from Friday 10/26, Lecture 28). Supose that $F_1, F_2, \ldots$ is a sequence of distribution functions with corresponding characteristic functions $\phi_1, \phi_2, \ldots$.

(a) If $F_n(x) \to F(x)$ for some distribution function F with characteristic function ϕ (at x where F is continuous), then $\phi_n(t) \to \phi(t)$ for all t.

(b) Conversely, if $\phi(t) = \lim_{n\to\infty} \phi_n(t)$ exists and $\phi(t)$ is continuous at $t = 0$, then ϕ is the characteristic function of some distribution function F, and $F_n \to F$.

Proof. See Kingman and Taylor (1966). □

7.2 Inequalities (8.6 of Pesaran)

Inequalities

- Probabilities
  -

    Lemma 40. Markov's Inequality (Grimmett and Stirzaker p. 311, 319)): For $a > 0$,

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

Proof. Note that $a \cdot \mathbf{1}_{\{|X| \geq a\}} \leq |X|$, where $\mathbf{1}$ is the indicator function. Dividing both sides by a and taking expectations yields the result. $\square$

–

Theorem 41. Chebyshev's Inequality: (probability p. 319) Let X be an (integrable) random variable with finite expected value μ and finite nonzero variance σ^2. Then for any real number $k > 0$

$$\Pr\big(|X - \mu| \geq k\sigma \big) \leq \frac{1}{k^2}$$

(Can be used to demonstrate consistency of estimators: if we can show that as $T \to \infty$ $\mathrm{Var}(X) = \sigma^2 \to 0$, then this implies $\Pr\big(|X - \mu| \geq k\sigma \big) \to 0$ as $T \to \infty$, showing consistency.)

–

Theorem 42. Chernoff For $x \geq 0$, $a > 0$, $\forall\ t > 0$,

$$\Pr(X \geq a) = \Pr(e^{tx} \geq e^{ta}) \leq \frac{\mathbb{E}(e^{tx})}{e^{ta}}$$

- **Moments**

  –

  Theorem 43. Cauchy-Schwarz. (and **Bunyakovsky**)

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

  – **Krylov**

  –

  Theorem 44. Jensen's (Grimmett and Stirzaker p.181, 349) If u is convex and $\mathbb{E}X < \infty$,

  $$\mathbb{E}(u(X)) \geq u(\mathbb{E}(X))$$

  –

  Theorem 45. Holder (Grimmett and Stirzaker p. p. 143, 319) Generalization of Cauchy-Schwarz. For $p, q > 1$ satisfying $1/p + 1/q = 1$ we have

  $$\mathbb{E}(|XY|) \leq (\mathbb{E}(|X^p|))^{1/p}(\mathbb{E}(|X^q|))^{1/q}$$

  –

  Theorem 46. Minkowski (Grimmett and Stirzaker p. p. 143) For $p \geq 1$,

  $$[\mathbb{E}(|X + Y|^p)]^{1/p} \leq (\mathbb{E}|X^p|)^{1/p} + (\mathbb{E}|Y^p|)^{1/p}$$

  – Useful for showing lower order moments are finite (e.g. finite variance implies finite mean).

  Lemma 47. Lyapunov's Inequality (Grimmett and Stirzaker p. 143). For $0 < r \leq s < \infty$,

  $$\mathbb{E}(|X|^r)^{1/r} \leq \mathbb{E}(|X|^s)^{1/s}$$

Monotone convergence theorem.

Dominated convergence theorem.

7.3 Modes of Convergence (7.2 of Grimmet and Strikazer, 8.2 and 8.4 of Pesaran)

Let $\{X_n\} = \{X_1, X_2, \ldots\}$ and X be random variables defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

Definition 7.6. Convergence in probability. $\{X_n\}$ is said to **converge in probability** to X if

- Grimmett and Strizaker definition:

$$\lim_{n\to\infty} \Pr(|X_n - X| > \epsilon) = 0, \text{ for every } \epsilon > 0$$

- Pesaran definition:

$$\lim_{n\to\infty} \Pr(|X_n - X| < \epsilon) = 1, \text{ for every } \epsilon > 0$$

This mode of convergence is also often denoted by $X_n \xrightarrow{p} X$ and when X is a fixed constant it is referred to as the **probability limit of** X_n, written as $Plim(X_n) = x$, as $n \to \infty$.

The above concept is readily extended to multivariate cases where $\{\boldsymbol{X}_n, n = 1, 2, \ldots\}$ denote m-dimensional vectors of random variables. Then the condition is

$$\lim_{n\to\infty} \Pr(\|\boldsymbol{X}_n - \boldsymbol{X}\| < \epsilon) = 1, \text{ for every } \epsilon > 0$$

where $\|\cdot\|$ denotes an appropriate norm (say ℓ_2). Convergence in probability is often referred to as "weak convergence" (in contrast to convergence with probability 1, below).

Definition 7.7. Convergence with probability 1 or almost surely. The sequence of random variables $\{X_n\}$ is said to **converge with probability 1** (or **almost surely**) to X if

- (505A class notes definition)

$$\Pr\left(\{\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\}\right) = 1$$

  (Note: pointwise convergence can hardly ever be shown here and is not useful.)

- Grimmett and Strikazer textbook definition:

$$\Pr\left(\{\omega \in \Omega : X_n(\omega) \to X(\omega) \text{ as } n \to \infty\}\right) = 1$$

- Pesaran textbook definition:

$$\Pr\left(\lim_{n\to\infty} X_n = X\right) = 1$$

This is often written as $X_n \xrightarrow{w.p.1} X$ or $X_n \xrightarrow{a.s.} X$. An equivalent condition for convergence with probability 1 is given by

$$\lim_{n\to\infty} \Pr(|X_m - X| < \epsilon, \text{ for all } m \geq= n) = 1, \text{ for every } \epsilon > 0$$

which shows that convergence in probability is a special case of convergence with probability 1 (obtained by setting $m = n$). Convergence with probability 1 is stronger than convergence in probability and is often referred to as "strong convergence."

Definition 7.8. Convergence in r-th mean. $X_n \to X$ **in rth mean** where $r \geq 1$ if $\mathbb{E}|X_n^r| < \infty$ for all n and

$$\lim_{n\to\infty} \mathbb{E}(|X_n - X|^r) = 0$$

Convergence in rth mean is often written $X_n \xrightarrow{r} X$.

Definition 7.9. Convergence in Distribution. Let $X_1, X_2, \ldots$ have distribution functions $F_1(\cdot), F_2(\cdot), \ldots$ respectively. Then X_n is said to **converge in distribution to** X if

$$\lim_{n\to\infty} \Pr(X_n \leq u) = \Pr(X \leq u)$$

for all u at which $F_X(x) = \Pr(X \leq x)$ is continuous. This can also be written

$$\lim_{n\to\infty} F_n(u) = F(u)$$

for all u at which F is continuous. Convergence in distribution is usually denoted by $X_n \xrightarrow{d} X$, $X_n \xrightarrow{L} X$, or $F_n \implies F$. By the Continuity Theorem (section 7.1), this is equivalent to

$$\lim_{n\to\infty} \phi_{X_n}(t) = \phi_X(t), \quad t \in \mathbb{R}$$

Theorem 48. (Theorem 7.2.3, Grimmett and Stirzaker.) The following implications hold:

- $(X_n \xrightarrow{a.s.} X) \implies (X_n \xrightarrow{p} X)$
- $(X_n \xrightarrow{r} X) \implies (X_n \xrightarrow{p} X)$ for any $r \geq 1$
- $(X_n \xrightarrow{p} X) \implies (X_n \xrightarrow{d} X)$

Also, if $r > s \geq 1$, then $(X_n \xrightarrow{r} X) \implies (X_n \xrightarrow{s} X)$. No other implications hold in general.

Theorem 49. Some exceptions (Theorem 7.2.4).

- If $X_n \xrightarrow{d} c$ where c is constant, then $X_n \xrightarrow{p} c$.
- If $X_n \xrightarrow{p} X$ and $\Pr(|X_n| \leq k) = 0$ for all n and some k, then $X_n \xrightarrow{r} X$ for all $r \geq 1$.
- If $P_n(\epsilon) = \Pr(|X_n - X| > \epsilon)$ satisfies $\sum_n P_n(\epsilon) < \infty$ for all $\epsilon > 0$, then $X_n \xrightarrow{a.s.} X$.

Proof. (Part (c).) Let $A_n(\epsilon) = \{|X_n - X| > \epsilon\}$ (so that $P_n(\epsilon) = \Pr[A_n(\epsilon)]$], and let $B_m(\epsilon) = \bigcup_{n \geq m} A_n(\epsilon)$. Then

$$\Pr(B_m(\epsilon)) \leq \sum_{n=m}^{\infty} \Pr(A_n(\epsilon))$$

so $\lim_{m\to\infty} \Pr(B_m(\epsilon)) = 0$ whenever $\sum_n \Pr(A_n(\epsilon)) < \infty$. See also Lemma 51 part (b). $\square$

7.4 More on convergence (7.2 of Grimmet and Strikazer)

Other theorems to include: Fatou's Lemma, Fubini's Theorem, Kolmogorov's Maximal Inequality, Kolmogorov Three-Series Test, Lindeberg Feller Central Limit Theorem, **this and more at beginning of Mike's 505A qual solutions.**

Definition 7.10. Cauchy Convergence. We say that the sequence $\{X_n : n \geq 1\}$ of random variables on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is **almost surely Cauchy convergent** if

$$\Pr\left(\{\omega \in \Omega : X_m(\omega) - X_n(\omega) \to 0 \text{ as } m, n \to \infty\}\right) = 1$$

That is, the set of points ω of the sample space for which the real sequence $\{X_n(\omega) : n \geq 1\}$ is Cauchy convergent is an event having probability 1.

Lemma 50. (Lemma 7.2.6 from Grimmett and Stirzaker)

(a) If $r > s \geq 1$ and $X_n \xrightarrow{r} X$, then $X_n \xrightarrow{s} X$.

(b) If $X_n \xrightarrow{1} X$ then $X_n \xrightarrow{p} X$.

The converse assertions fail in general.

Proof. (a) Using Lyapunov's Inequality (Lemma 47), if $r > s \geq 1$

$$\left[\mathbb{E}(|X_n - X|^s)\right]^{1/s} \leq \left[\mathbb{E}(|X_n - X|^r)\right]^{1/r}$$

Therefore if $X_n \xrightarrow{r} X$ (meaning $\lim_{n\to\infty} \mathbb{E}(|X_n - X|^r) = 0$), (then $\lim_{n\to\infty} \mathbb{E}(|X_n - X|^s) = 0$, so $X_n \xrightarrow{s} X$. We show the converse fails by counterexample:

$$X_n = \begin{cases} n & \text{with probability } n^{(-1/2)(r+s)} \\ 0 & \text{with probability } 1 - n^{(-1/2)(r+s)} \end{cases}$$

Then $\mathbb{E}|X_n^s| = n^{(1/2)(s-r)} \to 0$ and $\mathbb{E}|X_n^r| = n^{(1/2)(r-s)} \to \infty$.

(b) By Markov's Inequality (Lemma 40),

$$\Pr(|X_n - X| > \epsilon) \leq \frac{\mathbb{E}|X_n - X|}{\epsilon} \quad \text{for all } \epsilon > 0$$

Therefore if $X_n \xrightarrow{1} X$; that is, $\lim_{n\to\infty} \mathbb{E}(|X_n - X|) = 0$, then $\lim_{n\to\infty} \Pr(|X_n - X| > \epsilon) = 0$ for every $\epsilon > 0$, so $X_n \xrightarrow{p} X$.

To see the converse fails, define an independent sequence $\{X_n\}$ by

$$X_n = \begin{cases} n^3 & \text{with probability } n^{-2} \\ 0 & \text{with probability } 1 - n^{-2} \end{cases}$$

Then $\Pr(|X| > \epsilon) = n^{-2}$ for all large n, and so $X_n \xrightarrow{p} 0$. However, $\mathbb{E}|X_n| = n \to \infty$.

□

Lemma 51. (Lemma 7.2.10, Grimmett and Stirzaker.) Let $A_n(\epsilon) = \{|X_n - X| > \epsilon\}$ and $B_m(\epsilon) = \cup_{n \geq m} A_n(\epsilon)$. Then:

(a) $X_n \xrightarrow{a.s.} X$ if and only if $\Pr(B_m(\epsilon)) \to 0$ as $m \to \infty$ for all $\epsilon > 0$.

(b) $X_n \xrightarrow{a.s.} X$ if $\sum_n \Pr(A_n(\epsilon)) < \infty$ for all $\epsilon > 0$.

(c) If $X_n \xrightarrow{a.s.} X$ then $X_n \xrightarrow{p} X$, but the converse fails in general.

Proof. (a)

(b) As for Theorem 49 part (c).

(c) To see the converse fails, define an independent sequence $\{X_n\}$ by

$$X_n = \begin{cases} 1 & \text{with probability } n^{-1} \\ 0 & \text{with probability } 1 - n^{-1} \end{cases}$$

Clearly $X_n \xrightarrow{p} 0$. However, if $0 < \epsilon < 1$,

$$\Pr(B_m(\epsilon)) = 1 - \lim_{r \to \infty} \Pr(X_n = 0 \text{ for all } n \text{ such that } m \leq n \leq r) \text{ (by Lemma 1.3.5)}$$

$$= 1 - \left(1 - \frac{1}{m}\right)\left(1 - \frac{1}{m+1}\right) \cdots \text{ (by independence)}$$

$$= 1 - \lim_{M \to \infty} \left(\frac{m-1}{m} \cdot \frac{m}{m+1} \cdot \frac{m+1}{m+2} \cdots \frac{M}{M+1} \right)$$

$$= 1 - \lim_{M \to \infty} \frac{m-1}{M+1} = 1$$

and so $\{X_n\}$ does not converge almost surely.

□

Lemma 52. (Lemma 7.2.12, Grimmett and Stirzaker.) There exist sequences which

(a) converge almost surely but not in mean,

(b) converge in mean but not almost surely.

Proof. (a) As for Lemma 50 part (b).

□

Theorem 53. (Theorem 7.2.13, Grimmett and Stirzaker.) If $X_n \xrightarrow{p} X$, there exists a non-random increasing sequence of integers $n_1, n_2, \ldots$ such that $X_{n_i} \xrightarrow{a.s.} X$ as $i \to \infty$.

Theorem 54. Skorokhod's representation theorem (Theorem 7.2.14, Grimmett and Stirzaker). If $\{X_n\}$ and X with distribution functions $\{F_n\}$ and F are such that $X_n \xrightarrow{d} X$ (or equivalently, $F_n \to F$) as $n \to \infty$, then there exists a probability space $(\Omega', \mathcal{F}', \mathbb{P}')$ and random variables $\{Y_n\}$ and Y mapping Ω' into $\mathbb{R}$ such that

(a) $\{Y_n\}$ and Y have distribution functions $\{F_n\}$ and F

(b) $Y_n \xrightarrow{a.s.} Y$ as $n \to \infty$

Therefore, although X_n may fail to converge to X in any mode other than in distribution, there exists a sequence $\{Y_n\}$ such that Y_n is distributed identically to X_n for every n, which converges almost surely to a copy of X.

Theorem 55. (Theorem 7.2.18, Grimmett and Stirzaker.) If $X_n \xrightarrow{d} X$ and $g : \mathbb{R} \to \mathbb{R}$ is continuous, then $g(X_n) \xrightarrow{d} g(X)$.

Theorem 56. (Theorem 7.2.19, Grimmett and Stirzaker; same as Portmanteau Theorem?) The following three statements are equivalent:

(a) $X_n \xrightarrow{d} X$

(b) $\mathbb{E}[g(X_n)] \to \mathbb{E}[g(X)]$ for all bounded continuous functions g.

(c) $\mathbb{E}[g(X_n)] \to \mathbb{E}[g(X)]$ for all functions g of the form $g(x) = f(x)\mathbf{1}_{[a,b]}(x)$ where f is continuous on $[a, b]$ and a and b are points of continuity of the distribution function of the random variable X.

Theorem 57. (Grimmett and Stirzaker Theorem 7.3.9.)

(a) If $X_n \xrightarrow{a.s.} X$ and $Y_n \xrightarrow{a.s.} Y$ then $X_n + Y_n \xrightarrow{a.s.} X + Y$.

(b) If $X_n \xrightarrow{r} X$ and $Y_n \xrightarrow{r} Y$ then $X_n + Y_n \xrightarrow{r} X + Y$.

(c) If $X_n \xrightarrow{p} X$ and $Y_n \xrightarrow{p} Y$ then $X_n + Y_n \xrightarrow{p} X + Y$.

(d) It is not in general true that $X_n + Y_n \xrightarrow{d} X + Y$ whenever $X_n \xrightarrow{d} X$ and $Y_n \xrightarrow{d} Y$.

Theorem 58. Borel-Cantelli lemmas (Grimmett and Stirzaker Theorem 7.3.10.) Let $\{A_n\}$ be an infinite sequence of events from some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let $A = \bigcap_n \bigcup_{m=n}^{\infty} A_m \ = \limsup_{n\to\infty} A_n = \{A_n \text{ i.o.}\}$ be the event that infinitely many of the A_n occur. Then:

(a) $\Pr(A) = 0$ if $\sum_n \Pr(A_n) < \infty$

(b) $\Pr(A) = 1$ if $\sum_n \Pr(A_n) = \infty$ and $A_1, A_2, \ldots$ are independent events.

Proof. (a) We have that $A \subseteq \bigcup_{m=n}^{\infty} A_n$ for all n, so

$$\Pr(A) \leq \sum_{m=n}^{\infty} \Pr(A_m) \to 0 \text{ as } n \to \infty$$

whenever $\sum_n \Pr(A_n) < \infty$.

(b) One can confirm that

$$A^c = \bigcup_n \bigcap_{m=n}^{\infty} A_m^c$$

But

$$\Pr\left(\bigcap_{m=n}^{\infty} A_m^c\right) = \lim_{r\to\infty} \Pr\left(\bigcap_{m=n}^{r} A_m^c\right) = \prod_{m=n}^{\infty}[1-\Pr(A_m)] \text{ (by independence) } \leq \prod_{m=n}^{\infty} \exp(-\Pr(A_m))$$

$$= \exp\left(-\sum_{m=n}^{\infty} \Pr(A_m)\right) = 0$$

whenever $\sum_n \Pr(A_n) = \infty$, where the fourth step follows since $1 - x \leq e^{-x}$ if $x \geq 0$. Thus

$$\Pr(A^c) = \lim_{n\to\infty} \Pr\left(\bigcap_{m=n}^{\infty} A_m^c\right) = 0$$

so $\Pr(A) = 1$.

□

Theorem 59. Kolmogorov's Two-Series Theorem. Let $X_1, X_2, \ldots$ be independent random variables with $\mathbb{E}(X_n) = \mu_n$ and $\mathrm{Var}(X_n) = \sigma_n^2$ such that $\sum_{n=1}^{\infty} \mu_n < \infty$ and $\sum_{n=1}^{\infty} \sigma_n^2 < \infty$. Then $\sum_{n=1}^{\infty} X_n$ converges in $\mathbb{R}$ almost surely.

Proof. Available on wikipedia, https://en.wikipedia.org/wiki/Kolmogorov%27s_two-series_theorem.

□

7.4.1 Slutsky's Convergence Theorems (8.4.1 of Pesaran, 7.3 of Grimmett and Stirzaker)

Theorem 60. Theorem 6 of Pesaran, Section 8.4.1, p. 173. Let $\{x_t, y_t\}, t = 1, 2, \ldots$ be a sequence of pairs of random variables with $y_t \xrightarrow{d} y$ and $|y_t - x_t| \xrightarrow{p} 0$. Then $x_t \xrightarrow{d} y$.

Theorem 61. Theorem 7 in Pesaran, on p.318 (section 7.3) of Grimmett and Stirzaker. (Section 8.4.1, p. 174) If $x_t \xrightarrow{d} x$ and $y_t \xrightarrow{p} c$ where c is a finite constant, then

(i) $x_t + y_t \xrightarrow{d} x + c$

(ii) $y_t x_t \xrightarrow{d} cx$

(iii) $x_t / y_t \xrightarrow{d} x/c$, if $c \neq 0$.

Theorem 62. on p.318 (section 7.3) of Grimmett and Stirzaker. Suppose that $X_n \xrightarrow{d} 0$ and $Y_n \xrightarrow{p} Y$, and let $g : \mathbb{R}^2 \to \mathbb{R}$ be such that $g(x, y)$ is a continuous function of y for all x, and $g(x, y)$ is continuous at $x = 0$ for all y. Then $g(X_n, Y_n) \xrightarrow{p} g(0, Y)$.

Theorem 63. Continuous Mapping Theorem (Theorem 9 of Pesaran, Section 8.4.1, p. 176: convergence properties of transformed sequences.) Suppose $\{\boldsymbol{x}_t\}$, $\{\boldsymbol{y}_t\}$, $\boldsymbol{x}$, and $\boldsymbol{y}$ are $m \times 1$ vectors of random variables on a probability space, and let $\boldsymbol{g}(\cdot)$ be a continuous vector-valued function. (Alternatively, suppose g has the set of discontinuity points D_g such that $\Pr(X \in D_g) = 0$.) Then

(i) $\boldsymbol{x}_t \xrightarrow{a.s.} x \implies \boldsymbol{g}(\boldsymbol{x}_t) \xrightarrow{a.s.} \boldsymbol{g}(\boldsymbol{x})$

(ii) $\boldsymbol{x}_t \xrightarrow{p} x \implies \boldsymbol{g}(\boldsymbol{x}_t) \xrightarrow{p} \boldsymbol{g}(\boldsymbol{x})$

(iii) $\boldsymbol{x}_t \xrightarrow{d} x \implies \boldsymbol{g}(\boldsymbol{x}_t) \xrightarrow{d} \boldsymbol{g}(\boldsymbol{x})$

(iv) $\boldsymbol{x}_t - \boldsymbol{y}_t \xrightarrow{p} \mathbf{0}$ and $\boldsymbol{y}_t \xrightarrow{d} \boldsymbol{y} \implies \boldsymbol{g}(\boldsymbol{x}_t) - \boldsymbol{g}(\boldsymbol{y}_t) \xrightarrow{d} \mathbf{0}(\boldsymbol{x})$

Proof. See Serfling (1980) or Rao (1973). □

7.5 Stochastic orders $\mathcal{O}_p(\cdot)$ and $o_p(\cdot)$ (Pesaran 8.5)

Definition 7.11. (Pesaran 8.5 Definition 6.) Let $\{a_t\}$ be a sequence of positive numbers and $\{x_t\}$ be a sequence of random variables. Then

(i) $x_t = \mathcal{O}_p(a_t)$, or x_t/a_t is bounded in probability, if for every $\epsilon > 0$ there exist real numbers M_ϵ and N_ϵ such that

$$\Pr\left(\frac{|x_t|}{a_t} > M_\epsilon\right) < \epsilon, \quad \text{for } t > N_\epsilon$$

(ii) $x_t = o_p(a_t)$ if

$$\frac{\boldsymbol{x}_t}{a_t} \xrightarrow{p} 0$$

7.6 Laws of Large Numbers and Central Limit Theorems (Pesaran 8.6; Grimmett and Stirzaker 7.4, 7.5)

Theorem 64. Weak Law of Large Numbers (Khinchine) (Pesaran 8.6 Theorem 10, Grimmett and Stirzaker Theorem 7.4.7). Suppose that $\{X_k\}$ is a sequence of (i) IID random variables with (ii) constant means, i.e., $\mathbb{E}(X_k) = \mu < \infty$. Then

$$\overline{X}_k = \frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow{p} \mu$$

Theorem 65. Weak Law of Large Numbers (Chebyshev) (Pesaran Section 8.6, p. 178, Theorem 11.) Let $\{X_k\}$ be a sequence of random variables. If (i) $\mathbb{E}(X_k) = \mu_k$, (ii) $\mathrm{Var}(X_k) = \sigma_k^2$, and (iii) $\mathrm{Cov}(X_k, X_j) = 0,\ k \neq j$, and (iv)

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} \sigma_k^2 < \infty$$

then we have $\overline{X}_n - \overline{\mu}_n \xrightarrow{p} 0$, where $\overline{\mu}_n = n^{-1} \sum_{k=1}^n \mu_k$.

Theorem 66. Strong Law of Large Numbers (Grimmett and Stirzakker Theorem 7.4.3). Let $\{X_k\}$ be a sequence of (i) independent (ii) identically distributed random variables with (iii) $\mathbb{E}(X_k) = \mu$ and (iv) $\mathbb{E}(X_k^2) < \infty$. Then

$$\frac{1}{n}\sum_{k=1}^n X_k \to \mu \text{ almost surely and in mean square.}$$

Theorem 67. Strong Law of Large Numbers (Grimmett and Stirzakker Theorem 7.5.1). Let $\{X_k\}$ be a sequence of (i) independent (ii) identically distributed random variables. Then if and only if (iii) $\mathbb{E}|X_k| < \infty$,

$$\frac{1}{n}\sum_{k=1}^n X_k \xrightarrow{a.s.} \mu$$

Theorem 68. Strong Law of Large Numbers 1 (Kolmogorov) (Pesaran 8.8 Theorem 12). Let $\{X_k\}$ be a sequence of (i) independent random variables with (ii) $\mathbb{E}(X_k) = \mu_k < \infty$ and (ii) $\mathrm{Var}(X_k) = \sigma_k^2$ such that (iii)

$$\sum_{k=1}^\infty \frac{\sigma_k^2}{k^2} < \infty$$

Then $\overline{X}_n - \overline{\mu}_n \xrightarrow{wp1} 0$. If the independence assumption (i) is replaced by a lack of correlation (i.e. $\mathrm{Cov}(X_k, X_j) = 0, k \neq j$), the convergence of $\overline{X}_n - \overline{\mu}_n$ with probability one requires the stronger condition

$$\sum_{k=1}^\infty \frac{\sigma_k^2 (\log k)^2}{k^2} < \infty$$

Theorem 69. Strong Law of Large Numbers 2 (Pesaran 8.8 Theorem 13) Suppose that $X_1, X_2, \ldots$ are (i) independent random variables, and that (ii) $\mathbb{E}(X_k) = 0$, (iii) $\mathbb{E}(X_k^4) \leq M \; \forall \; k$ where M is an arbitrary positive constant. Then

$$\overline{X}_n = \frac{1}{n}\sum_{k=1}^n X_k \xrightarrow{a.s.} 0$$

Theorem 70. Central Limit Theorem (Grimmett and Stirzaker theorem 5.10.4.) Let $X_1, X_2, \ldots$ be a sequence of independent identically distributed random variables with finite mean μ and finite non-zero variance σ^2, and let $S_n = \sum_{i=1}^n X_i$. Then

$$\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \xrightarrow{d} \mathcal{N}(0, 1)$$

Theorem 71. (Grimmett and Stirzaker theorem 5.10.5.) Let $X_1, X_2, \ldots$ be independent random variables satisfying $\mathbb{E}(X_j) = 0$, $\mathrm{Var}(X_j) = \sigma_j^2$, $\mathbb{E}|X_j^3| < \infty$ such that

$$\lim_{n\to\infty} \frac{1}{\sigma(n)^3} \sum_{j=1}^n \mathbb{E}|X_j^3| = 0$$

where $\sigma(n)^2 = \mathrm{Var}\big(\sum_{j=1}^n X_j\big) = \sum_{j=1}^n \sigma_j^2$. Then

$$\frac{1}{\sigma(n)} \sum_{j=1}^n X_j \xrightarrow{d} \mathcal{N}(0,1)$$

Proof. See Loeve (1977, p. 287) and Grimmett and Stirzaker Problem 5.12.40. □

Lemma 72. Lindeberg's Condition: Let $\{X_k\}$ be a sequence of independent (not necessarily identically distributed) random variables with expectations μ_k and finite variances σ_k^2. Let $s_n^2 = \sum_{k=1}^n \sigma_k^2$. If such a sequence of independent random variables X_k satisfies the condition

$$\lim_{n\to\infty} \frac{1}{s_n^2} \sum_{k=1}^n \mathbb{E}\big[(X_k - \mu_k)^2) \cdot \mathbf{1}_{\{|X_k - \mu_k| > \epsilon s_n\}}\big] = 0$$

for all $\epsilon > 0$ then the central limit theorem holds; that is, the random variables

$$Z_n = \frac{1}{s_n} \sum_{k=1}^n (X_k - \mu_k)$$

converge in distrbution to $\mathcal{N}(0,1)$ as $n \to \infty$.

7.7 The case of dependent and heterogeneously distributed observations (Pesaran 8.8)

Theorem 73. Central limit theorem for martingale difference sequences (Pesaran 8.8 Theorem 28). Let $\{x_t\}$ be a martingale difference sequence with respect to the information set Ω_t. Let $\overline{\sigma}_T^2 = \mathrm{Var}(\sqrt{T}\overline{x}_T) = T^{-1}\sum_{t=1}^T \sigma_t^2$. If $\mathbb{E}(|x_t|^r) < K < \infty$, $r > 2$ and for all t, and

$$\frac{1}{T} \sum_{t=1}^T x_t^2 - \overline{\sigma}_t^2 \xrightarrow{p} 0$$

then $\sqrt{T}\overline{x}_T / \overline{\sigma}_T \xrightarrow{d} \mathcal{N}(0,1)$.

7.8 Worked Examples from Math 505A Midterm 2

(1) (a) Let X_k, $k \geq 1$, be i.i.d. random variables with mean 1 and variance 1. Show that the limit

$$\lim_{n\to\infty} \frac{\sum_{k=1}^n X_k}{\sum_{k=1}^n X_k^2}$$

exists in an appropriate sense, and identify the limit.

(b) Let $(X_j)_{j\geq 1}$ be i.i.d. uniform on $(-1, 1)$. Let

$$Y_n = \frac{\sum_{j=1}^n X_j}{\sum_{j=1}^n X_j^2 + \sum_{j=1}^n X_j^3}$$

Prove that $\lim_{n\to\infty} \sqrt{n} Y_n$ exists in an appropriate sense, and identify the limit.

Solution.

(a)

$$\lim_{n\to\infty} \frac{\sum_{k=1}^n X_k}{\sum_{k=1}^n X_k^2} = \lim_{n\to\infty} \frac{n^{-1}\sum_{k=1}^n X_k}{n^{-1}\sum_{k=1}^n X_k^2}$$

Since $X_1, X_2, \ldots$ are i.i.d., $E(X_1^2) = \text{Var}(X_1) + (\mathbb{E}(X_1))^2 = 2 < \infty$, we have

$$n^{-1} \sum_{k=1}^n X_k \xrightarrow{a.s.} \mathbb{E}(X_1) = 1 \text{ as } n \to \infty$$

by Theorem 66 (Strong Law of Large Numbers). Also, $X_1^2, X_2^2, \ldots$ are clearly identically distributed, and are independent by Theorem 4.2.3 ("If X and Y are independent, then so are $g(X)$ and $g(Y)$."). It is clear also that $\mathbb{E}(|X_1^2|) = \mathbb{E}(X_1^2) = \text{Var}(X_1) + \mathbb{E}(X_1)^2 = 1 + 1 = 2 < \infty$. Therefore by Theorem 67 (Strong Law of Large Numbers),

$$n^{-1} \sum_{k=1}^n X_k^2 \xrightarrow{a.s.} \mathbb{E}(X_1^2) = 2 \text{ as } n \to \infty$$

(From here I had two different ways of finishing the problem.)

- Because we have almost sure convergence in the numerator and denominator, by the Continuous Mapping Theorem (Theorem 63),

$$\lim_{n\to\infty} \frac{n^{-1}\sum_{k=1}^n X_k}{n^{-1}\sum_{k=1}^n X_k^2} = \frac{\lim_{n\to\infty} n^{-1}\sum_{k=1}^n X_k}{\lim_{n\to\infty} n^{-1}\sum_{k=1}^n X_k^2} \xrightarrow{a.s.} \boxed{\frac{1}{2}}$$

- Then, using one of Slutsky's convergence theorems (Theorem 61: "If $x_t \xrightarrow{d} x$ and $y_t \xrightarrow{p} c$ where c is a finite constant, then $x_t/y_t \xrightarrow{d} x/c$, if $c \neq 0$."), we have

$$\frac{n^{-1}\sum_{k=1}^n X_k}{n^{-1}\sum_{k=1}^n X_k^2} \xrightarrow{d} \frac{\mathbb{E}(X_1)}{\mathbb{E}(X_1^2)} = \frac{\mathbb{E}(X_1)}{\text{Var}(X_1) + \mathbb{E}(X_1)^2} = \frac{1}{1+1} = \frac{1}{2}$$

  But then, by Theorem 49 (Theorem 7.2.4(a) in Grimmett and Stirzaker: "If $X_n \xrightarrow{d} c$ where c is constant, then $X_n \xrightarrow{p} c$."), we have $\frac{n^{-1}\sum_{k=1}^n X_k}{n^{-1}\sum_{k=1}^n X_k^2} \xrightarrow{p} 1/2$.

(b)

$$Y_n = \frac{\sum_{j=1}^n X_j}{\sum_{j=1}^n X_j^2 + \sum_{j=1}^n X_j^3} = \frac{n^{-1}\sum_{j=1}^n X_j}{n^{-1}\sum_{j=1}^n X_j^2 + n^{-1}\sum_{j=1}^n X_j^3}$$

Note that $\mathbb{E}(X_1) = 0, \mathbb{E}(X_1^2) = \text{Var}(X_1) + \mathbb{E}(X_1)^2 = (1 - -1)^2/12 + 0^2 = 1/3, \mathbb{E}(X_1^3) = (1/2)\int_{-1}^1 x^3 dx = 0$. (We derived the formulae for the first three moments of a uniform distribution on Homework 4 problem 2(2).)

$$\implies \sqrt{n} Y_n = \frac{\sqrt{1/3}\big(\sum_{j=1}^n X_j - n\mathbb{E}(X_1)\big)/\sqrt{n \cdot 1/3}}{n^{-1}\sum_{j=1}^n X_j^2 + n^{-1}\sum_{j=1}^n X_j^3}$$

By the Central Limit Theorem (Theorem 70),

$$\frac{\sum_{j=1}^n X_j - n\mathbb{E}(X_1)}{\sqrt{n \cdot 1/3}} \xrightarrow{d} \mathcal{N}(0,1)$$

By the Law of Large Numbers (Theorem 67), since $\mathbb{E}(|X_1^2|) = \mathbb{E}(X_1^2) = 1/3 < \infty$,

$$\frac{1}{n}\sum_{j=1}^n X_j^2 \xrightarrow{a.s.} \mathbb{E}(X_1^2) = 1/3$$

By the Law of Large Numbers (Theorem 67), since $\mathbb{E}(|X_1^3|) = (1/2)\int_{-1}^1 |x^3| dx = \int_0^1 x^3 dx = 1/4 < \infty$,

$$\frac{1}{n}\sum_{j=1}^n X_j^3 \xrightarrow{a.s.} \mathbb{E}(X_1^3) = 0$$

In the denominator, since we have almost sure convergence, the regular rules of calculus/real analysis apply. That is, using the above results,

$$n^{-1}\sum_{j=1}^n X_j^2 + n^{-1}\sum_{j=1}^n X_j^3 \xrightarrow{a.s.} 1/3$$

Therefore

$$\sqrt{n}Y_n = \frac{\sqrt{1/3}\big(\sum_{j=1}^n X_j - n\mathbb{E}(X_1)\big)/\sqrt{n\cdot 1/3}}{n^{-1}\sum_{j=1}^n X_j^2 + n^{-1}\sum_{j=1}^n X_j^3} \xrightarrow{d} \frac{\sqrt{1/3}}{1/3}\mathcal{N}(0,1) = \boxed{\mathcal{N}(0,3)}$$

(2) **Question:** Fix $p \in (0,1)$ and consider independent Poisson random variables $X_k, k \geq 1$ with

$$\mathbb{E}X_k = \frac{p^k}{k}$$

Verify that the sum $\sum_{k=1}^\infty kX_k$ converges with probability one and determine the distribution of the random variable $Y = \sum_{k=1}^\infty kX_k$.

Solution. Melike's solution (use for midterm): We have $\mathbb{E}[kX_k] = p^k$ and $\sum_{k=1}^\infty p^k = p/(1-p) < \infty$, and $\text{Var}(kX_k) = kp^k$ and

$$\sum_{k=1}^\infty kp^k = p\sum_{k=1}^\infty kp^{k-1} = p\frac{\mathrm{d}}{\mathrm{d}p}\sum_{k=1}^\infty p^k = p\frac{\mathrm{d}}{\mathrm{d}p}\frac{p}{1-p} = p\cdot\frac{(1-p)-p(-1)}{(1-p)^2} = \frac{p}{(1-p)^2} < \infty$$

Since the sequence $\{Y_k|\}_{k\geq 1}$ is independent, by Kolmogorov's Two Series Theorem (Theorem 59: "Let $X_1, X_2, \ldots$ be independent random variables with $\mathbb{E}(X_n) = \mu_n$ and $\text{Var}(X_n) = \sigma_n^2$ such that $\sum_{n=1}^\infty \mu_n < \infty$ and $\sum_{n=1}^\infty \sigma_n^2 < \infty$. Then $\sum_{n=1}^\infty X_n$ converges in $\mathbb{R}$ almost surely."), we conclude that $\sum_{k=1}^\infty kX_k$ converges almost surely.

To find the distribution of Y, let X be a Poisson random variable and consider its probability generating function:

$$G_X(s) = \mathbb{E}(s^X) = \sum_{k=0}^\infty s^k e^{-\lambda}\frac{\lambda^k}{k!} = e^{-\lambda}\sum_{k=0}^\infty \frac{(\lambda s)^k}{k!} = e^{-\lambda}e^{\lambda s} = e^{\lambda(s-1)}$$

So $\mathbb{E}(s^{X_k}) = \exp\left(\frac{p^k}{k}(s-1)\right)$ and $\mathbb{E}(s^{kX_k}) = \mathbb{E}[(s^k)^{X_k}] = \exp\left(\frac{p^k}{k}(s^k-1)\right)$. Then define $Y_n = \sum_{k=1}^n kX_k$ and consider

$$G_{Y_n}(s) = \mathbb{E}(s^{Y_n}) = \mathbb{E}\Big(\prod_{k=1}^{n} s^{kX_k}\Big) = \prod_{k=1}^{n} \mathbb{E}\big(s^{kX_k}\big) = \prod_{k=1}^{n} \exp\Big(\frac{p^k}{k}(s^k - 1)\Big) = \exp\Big(\sum_{k=1}^{n} \frac{p^k}{k}(s^k - 1)\Big)$$

$$= \exp\Big(\sum_{k=1}^{n} \frac{(ps)^k}{k} - \sum_{k=1}^{n} \frac{p^k}{k}\Big)$$

Now, by taking limits as $n \to \infty$ (since we are allowed to take limit inside of expectation here), we get

$$G_Y(s) = \mathbb{E}(s^Y) = \exp\Big(\sum_{k=1}^{\infty} \frac{(ps)^k}{k} - \sum_{k=1}^{\infty} \frac{p^k}{k}\Big) = \exp\Big(\int \sum_{k=1}^{\infty} (ps)^{k-1} dp - \int \sum_{k=1}^{\infty} p^{k-1} dp\Big)$$

$$= \exp\Big(\int \frac{1}{1-ps} dp - \int \frac{1}{1-p} dp\Big) = \exp(-\log(1-ps) + \log(1-p)), \quad -1 \le ps < 1 \text{ and } -1 \le p < 1$$

$$= \frac{1-p}{1-ps}, \quad -1 \le ps < 1$$

Since we know $\Pr(X = k) = \frac{G_X^{(k)}(0)}{k!}$, we have

$$G_Y(s) = \frac{1-p}{1-sp},\ G'(s) = \frac{p(1-p)}{(1-sp)^2},\ G''(s) = \frac{2p^2(1-p)}{(1-sp)^3},\ G^{(3)}(s) = \frac{3 \cdot 2p^3(1-p)}{(1-sp)^3}, \dots$$

$$G^{(k)}(s) = \frac{k! p^k (1-p)}{(1-sp)^k} \text{ for } k = 0, 1, 2, \dots$$

So we have

$$\Pr(Y = k) = (1-p)p^k, \quad k = 0, 1, 2, \dots$$

which means $Y \sim G_1(p) - 1$.

(3) (a) Consider the sequence $\{X_k, k \ge 1\}$ of random variables such that X_1 is uniform on $(0,1)$ and, given X_k, the distribution of X_{k+1} is uniform on $(0, CX_k)$, where $\sqrt{3} < C < 2$.

(i) Show that $\lim_{x\to\infty} X_n = 0$ in ℓ_1 and with probability one, but not in ℓ_2.

(ii) Investigate the same questions for all other values of $C > 0$.

(b) Let $a > 0$, let $X_n, n \ge 1$ be i.i.d. random variables that are uniform on $(0, a)$, and let $Y_n = \prod_{k=1}^{n} X_k$. Determine, with a proof, all values of a for which $\lim_{n\to\infty} Y_n = 0$ with probability one.

Solution.

(a) (i) We have that $X_{n+1} \mid X_n \sim U(0, CX_n)$. Therefore

$$\mathbb{E}(X_{n+1}^r \mid X_n) = \frac{1}{CX_n} \int_0^{CX_n} x^r dx = \frac{1}{CX_n} \cdot \frac{x^{r+1}}{r+1}\Big|_0^{CX_n} = \frac{C^r X_n^r}{r+1}$$

$$\implies \mathbb{E}(X_{n+1}^r) = \mathbb{E}[\mathbb{E}(X_{n+1}^r \mid X_n)] = \frac{C^r}{r+1} \cdot \mathbb{E}(X_n^r)$$

Note that $E(X_1^r) = \int_0^1 x^r dr = 1/(r+1)$. Therefore

$$\mathbb{E}(X_{n+1}^r) = \frac{C^r}{r+1} \cdot \mathbb{E}(X_n^r) = \left(\frac{C^r}{r+1} \right)^n \cdot \mathbb{E}(X_1^r) = \left(\frac{C^r}{r+1} \right)^n \cdot \frac{1}{r+1}$$

We would like to show that $X_n \xrightarrow{w.p.1} 0$ and that $X_n \xrightarrow{1} 0$, but that the same result does not follow for the ℓ_2 norm.

- **Convergence with probability one:** We seek to show that $\Pr\big(\{\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = 0\}\big) = 1$. By Markov's Inequality (Lemma 40), we have

$$\Pr(|X_n| \geq a) \leq \frac{\mathbb{E}(X_n)}{a} \quad \forall\, a > 0$$

$$\iff \Pr(|X_n| \geq a) \leq \left(\frac{C^1}{1+1} \right)^{n-1} \cdot \frac{1}{1+1} \cdot \frac{1}{a} = \left(\frac{C}{2} \right)^{n-1} \cdot \frac{1}{2a} \quad \forall\, a > 0$$

  Since $\sqrt{3} < C < 2$, $\sqrt{3}/2 < C/2 < 1$. Since $X_n \in [0, CX_{n-1}]$, $X_n \geq 0$, so $|X_n| = X_n$. Therefore we have

$$\Pr(\lim_{n\to\infty} |X_n| \geq a) = \Pr(\lim_{n\to\infty} X_n \geq a) \leq \lim_{n\to\infty} \left(\frac{C}{2} \right)^{n-1} \cdot \frac{1}{2a} = 0 \quad \forall\, a > 0$$

  Since $|X_n| \geq 0$, this implies that $\Pr(\lim_{n\to\infty} X_n = 0) = \Pr\big(\{\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = 0\}\big) = 1$, so by the Borel-Cantelli Lemma (Theorem 58), X_n converges to 0 with probability 1.

- **Convergence in ℓ_1 norm:** We seek to show that $\lim_{n\to\infty} \mathbb{E}(|X_n|) = 0$. Since $X_n \in [0, CX_{n-1}]$, $X_n \geq 0$, so $|X_n| = X_n$. Therefore

$$\lim_{n\to\infty} \mathbb{E}(|X_n|) = \lim_{n\to\infty} \mathbb{E}(X_n) = \lim_{n\to\infty} \left(\frac{C}{2} \right)^{n-1} \cdot \frac{1}{2}$$

  Since $\sqrt{3} < C < 2$, $\sqrt{3}/2 < C/2 < 1$, so $C/2 < 1$. Therefore we have

$$\lim_{n\to\infty} \mathbb{E}(|X_n|) = \lim_{n\to\infty} \left(\frac{C}{2} \right)^{n-1} \cdot \frac{1}{2} = 0$$

  so X_n converges to 0 in 1st mean.

- **Convergence in ℓ_2 norm:** We seek to show that $\lim_{n\to\infty} \mathbb{E}(|X_n|^2) \neq 0$. We have

$$\lim_{n\to\infty} \mathbb{E}(|X_n|^2) = \lim_{n\to\infty} \mathbb{E}(X_n^2) = \lim_{n\to\infty} \left(\frac{C^2}{3} \right)^{n-1} \cdot \frac{1}{3}$$

  Since $\sqrt{3} < C < 2$, $3/3 < C^2/3 < 4/3$, so $C^2/3 > 1$. Therefore we have

$$\lim_{n\to\infty} \mathbb{E}(|X_n|^2) = \lim_{n\to\infty} \left(\frac{C^2}{3} \right)^{n-1} \cdot \frac{1}{3} = \infty \neq 0$$

  so X_n does not converge to 0 in 2nd mean.

(ii) From the above, it is clear that for convergence with probability one or in 1st mean we require $0 < C/2 < 1$ and for convergence in second mean we require $0 < C^2/3 < 1$. For $0 < C < \sqrt{3}$, we see that X_n would converge to zero in 2nd mean since this would imply that $0 < C^2/3 < 1$. It would also still converge to 0 in 1st mean (and with probability 1) since we would have $(0 < C/2 < \sqrt{3}/2 < 1$.
For $C = \sqrt{3}$, X_n would still converge to 0 with probability one and in 1st mean for the same reasons. However, it would not converge in 2nd mean because we would have

$$\lim_{n\to\infty} \mathbb{E}(|X_n|^2) = \lim_{n\to\infty} \left(\frac{\sqrt{3}^2}{3} \right)^{n-1} \cdot \frac{1}{3} = \frac{1}{3} \neq 0$$

For $C \geq 2$, it would diverge in all three cases, since in this case $C/2 \geq 2/2 = 1$ and $C^2/3 \geq 4/3 > 1$.

(b) **Probably won't be on midterm.** Note that

$$\lim_{n\to\infty} Y_n = \lim_{n\to\infty} \prod_{k=1}^{n} X_k = 0 \iff \log(Y_n) = \log\left(\prod_{k=1}^{n} X_k\right) = \sum_{k=1}^{n} \log(X_k) \to -\infty$$

Note that

$$\mathbb{E}[\log(Y_n)] = \mathbb{E}\left(\sum_{k=1}^{n} \log(X_k)\right) = \sum_{k=1}^{n} \mathbb{E}[\log(X_k)] = \sum_{k=1}^{n} \mathbb{E}[\log(X_1)] = \sum_{k=1}^{n} \int_0^a (\log(x)/a)dx$$

$$= \sum_{k=1}^{n} \frac{1}{a} \left[x \log x - x \right]_0^a = \sum_{k=1}^{n} \frac{a \log a - a}{a} = \sum_{k=1}^{n} (\log(a) - 1) = n(\log(a) - 1)$$

As $n \to \infty$ we have

$$\mathbb{E}[\log(Y_n)] = \begin{cases} -\infty & a < e \\ 0 & a = e \\ \infty & a > e \end{cases}$$

Since $\mathbb{E}[\log(Y_n)] \to \infty$ for $a < e$, we have $\lim_{n\to\infty} Y_n = 0$ for $a < 3$. Therefore

$$\boxed{\lim_{n\to\infty} Y_n = \lim_{n\to\infty} \prod_{k=1}^{n} X_k = 0 \iff a < e.}$$

8 Linear Regression

These notes are based on my notes from *Time Series and Panel Data Econometrics* (1st edition) by M. Hashem Pesaran and coursework for Economics 613: Economic and Financial Time Series I at USC. I also borrowed from some other sources which I mention when I use them.

8.1 Chapter 1: Linear Regression

8.1.1 Preliminaries

Suppose the true model is $y_i = \alpha + \beta x_i + \epsilon_i$. Classical assumptions:

(i) $\mathbb{E}(\epsilon_i) = 0$

(ii) $\text{Var}(\epsilon_i \mid x_i = \sigma^2$ (constant)

(iii) $\text{Cov}(\epsilon_i, \epsilon_j) = 0$ if $i \neq j$

(iv) ϵ_i is uncorrelated to x_i, or $\mathbb{E}(\epsilon_i \mid x_j) = 0$ for all i, j.

8.1.2 Estimation

$$\hat{\beta} = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2} = \frac{\sum_{i=1}^n x_i y_i - n \overline{xy}}{\sum_{i=1}^n x_i^2 - n\overline{x}^2}$$

$$\hat{\alpha} = \overline{y} - \hat{\beta}\overline{x}$$

or

$$\hat{\beta} = \frac{\sum_{i=1}^n (x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^n (x_i - \overline{x})^2} = \frac{S_{XY}}{S_{XX}}$$

or

$$\hat{\beta} = r \frac{S_{YY}}{S_{XX}}$$

where r is the correlation coefficient.

Let

$$w_i = \frac{x_i - \overline{x}}{\sum_{i=1}^n (x_i - \overline{x})^2}$$

so that

$$\hat{\beta} = \sum_{i=1}^n w_i(y_i - \overline{y}) = \sum_{i=1}^n w_i y_i - \overline{y}\frac{\sum_{i=1}^n x_i - \overline{x}}{\sum_{i=1}^n (x_i - \overline{x})^2} = \sum_{i=1}^n w_i y_i$$

since $\sum_{i=1}^n x_i - \overline{x} = 0$. Then a simple expression for $\text{Var}(\hat{\beta})$ is

$$\text{Var}(\hat{\beta}) = \sum_{i=1}^n w_i^2 \text{Var}(y_i \mid x_i) = \sum_{i=1}^n w_i^2 \text{Var}(\epsilon \mid x_i) = \sigma^2 \sum_{i=1}^n w_i^2 = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \overline{x})^2} = \frac{\sigma^2}{S_{XX}}$$

We can estimate these quantities as follows:

$$\hat{\sigma}^2 = \frac{1}{n-2} \cdot \sum_{i=1}^n (y_i - \hat{\alpha} - \hat{\beta}x_i)^2$$

Note that

$$\hat{\sigma}^2 = \frac{1}{n-2}\sum_{t=1}^T (y_t - \hat{\alpha} - \hat{\beta}x_t)^2 = \frac{1}{n-2}\sum_{t=1}^T \left[(y_t - (\overline{y} - \hat{\beta}\overline{x}) - \hat{\beta}x_t)^2\right] = \frac{1}{n-2}\sum_{t=1}^T (y_t - \overline{y} - \hat{\beta}(x_t - \overline{x}))^2$$

$$= \frac{1}{n-2}\sum_{t=1}^T (y_t - \overline{y})^2 - 2\hat{\beta}(x_t - \overline{x})(y_t - \overline{y}) + \hat{\beta}^2(x_t - \overline{x})^2$$

In the case where there is no intercept, we have

$$\hat{\sigma}^2 = \frac{1}{T-1}\sum_{t=1}^T (y_t - \hat{\beta}x_t)^2 = \frac{1}{T-1}\sum_{t=1}^T \left(y_t^2 - 2r\frac{S_{YY}}{S_{XX}}x_t y_t + r^2\frac{S_{YY}^2}{S_{XX}^2}x_t^2 \right)$$

Also,

$$\widehat{\text{Var}}(\hat{\beta}) = \frac{\hat{\sigma}^2}{S_{XX}} = \frac{1}{n-2} \cdot \frac{\sum_{i=1}^n (y_i - \hat{\alpha} - \hat{\beta}x_i)^2}{\sum_{i=1}^n (x_i - \overline{x})^2}$$

Correlation coefficient:

$$r^2 = \frac{\left(\sum_{t=1}^T x_t y_t\right)^2}{\sum_{t=1}^T x_t^2 \sum_{t=1}^T y_t^2}$$

$$r = \frac{1}{T-1}\frac{S_{XY}}{\sqrt{S_{XX}S_{YY}}}$$

Remark. The formulas for the coefficients in univariate OLS can also be derived by considering (x, y) as a bivariate normal distribution and calculating the conditional expectation of y given x. (See Theorem (36).)

8.2 Chapter 2: Multiple Regression

General OLS:

$$\hat{\beta} = (\boldsymbol{X}'X)^{-1}X'y = (X'X)^{-1}X'(X\beta + u) = (X'X)^{-1}X'X\beta + (X'X)^{-1}X'u = \beta + (X'X)^{-1}X'u$$

$$\text{Var}(\hat{\beta}) = \text{Var}(\beta + (X'X)^{-1}X'u) = \text{Var}(\beta) + \text{Var}((X'X)^{-1}X'u) = 0 + \mathbb{E}[(X'X)^{-1}X'uu'X(X'X)^{-1}]$$

$$= \mathbb{E}\big[(X'X)^{-1}X'\mathbb{E}(uu' \mid X)X(X'X)^{-1}\big] = \sigma^2\mathbb{E}\big[(X'X)^{-1}X'I_TX(X'X)^{-1}\big] = \sigma^2\mathbb{E}\big[(X'X)^{-1}\big]$$

$$= \sigma^2(X'X)^{-1}$$

$$\hat{\sigma}^2 = \frac{\hat{\boldsymbol{u}}'\hat{\boldsymbol{u}}}{T-k}$$

8.3 Chapter 3: Hypothesis testing in regression

In this section, I borrow from C. Flinn's notes "Asymptotic Results for the Linear Regression Model," available online at http://www.econ.nyu.edu/user/flinnc/notes1.pdf.

Lemma 74.

$$\frac{1}{n} \cdot X'\epsilon \xrightarrow{p} 0$$

Proof. Note that $\mathbb{E}\frac{1}{n} \cdot X'\epsilon = 0$ for any n. Then we have

$$\text{Var}\left(\frac{1}{n} \cdot X'\epsilon\right) = \mathbb{E}\left(\frac{1}{n} \cdot X'\epsilon\right)^2 = n^{-2}\mathbb{E}(X'\epsilon\epsilon'X) = n^{-2}\mathbb{E}(\epsilon\epsilon')X'X = \frac{\sigma^2}{n}\frac{X'X}{n}$$

implying that $\lim_{n\to\infty} \text{Var}\left(\frac{1}{n} \cdot X'\epsilon\right) = 0$. Therefore the result follows from Chebyshev's Inequality (Theorem 41). □

Lemma 75. If ϵ is i.i.d. with $E(\epsilon_i) = 0$ and $\mathbb{E}(\epsilon_i^2) = \sigma^2$ for all i, the elements of the matrix X are uniformly bounded so that $|X_{ij}| < U$ for all i and j and for U finite, and $\lim_{n\to\infty} X'X/n = Q$ is finite and nonsingular, then

$$\frac{1}{\sqrt{n}}X'\epsilon \xrightarrow{d} \mathcal{N}(0, \sigma^2 Q)$$

Proof. If we have one regressor, then $n^{-1/2}\sum_{i=1}^n X_i\epsilon_i$ is a scalar. Let G_i be the cdf of $X_i\epsilon_i$. Let

$$S_n^2 = \sum_{i=1}^n \mathrm{Var}(X_i\epsilon_i) = \sigma^2 \sum_{i=1}^n X_i^2$$

In this scalar case, $Q = \lim_{n\to\infty} n^{-1}\sum_i X_i^2$. By the Lindberg-Feller Theorem, a necessary and sufficient condition for $Z_n \to \mathcal{N}(0\sigma^2 Q)$ is

$$\lim_{n\to\infty} \frac{1}{S_n^2}\sum_{i=1}^n \int_{|\omega|>\nu S_n} \omega^2 dG_i(\omega) = 0$$

for all $\nu > 0$. Now $G_i(\omega) = F(\omega/|X_i|)$. Then rewrite the above equation as

$$\lim_{n\to\infty} \frac{n}{S_n^2}\sum_{i=1}^n \frac{X_i^2}{n}\int_{|\omega/X_i|>\nu S_n/|X_i|} \left(\frac{\omega}{X_i}\right)^2 dF(\omega/|X_i|) = 0$$

Since $\lim_{n\to\infty} S_n^2 = \lim_{n\to\infty} n\sigma^2\sum_{i=1}^n X_i^2/n = n\sigma^2 Q$, we have $\lim_{n\to\infty} n/S_n^2 = (\sigma^2 Q)^{-1}$, which is a finite and nonzero scalar. Then we need to show

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n X_i^2 \delta_{i,n} = 0$$

where

$$\delta_{i,n} = \int_{|\omega/X_i|>\nu S_n/|X_i|} \left(\frac{\omega}{X_i}\right)^2 dF(\omega/|X_i|)$$

But $\lim_{n\to\infty}\delta_{i,n} = 0$ for all i and any fixed ν since $|X_i|$ is bounded while $\lim_{n\to\infty} X_n = \infty$, so the measure of the set $\{|\omega/X_i| > \nu S_n/|X_i|\}$ goes to 0 asymptotically. Since $\lim_{n\to\infty} n^{-1}\sum_i X_i^2$ is finite and $\lim_{n\to\infty}\delta_{i,n} = 0$ for all i, $\lim_{n\to\infty} n^{-1}\sum_i X_i^2\delta_{i,n} = 0$, so $\frac{1}{n}\cdot X'\epsilon \xrightarrow{p} 0$.

□

Theorem 76. Under the conditions of Lemma 75 (ϵ is i.i.d. with $E(\epsilon_i) = 0$ and $\mathbb{E}(\epsilon_i^2) = \sigma^2$ for all i, the elements of the matrix X are uniformly bounded so that $|X_{ij}| < U$ for all i and j and for U finite, and $\lim_{n\to\infty} X'X/n = Q$ is finite and nonsingular),

$$\sqrt{n}(\hat\beta - \beta) \xrightarrow{d} \mathcal{N}(0, \sigma^2 Q^{-1})$$

Proof.

$$\sqrt{n}(\hat\beta - \beta) = \left(\frac{X'X}{n}\right)^{-1}\frac{1}{\sqrt{n}}X'\epsilon$$

Since $\lim_{n\to\infty}(X'X/n)^{-1} = Q^{-1}$ and by Lemma 75

$$\frac{1}{\sqrt{n}} X' \epsilon \xrightarrow{d} \mathcal{N}(0, \sigma^2 Q)$$

then

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} \mathcal{N}(0, \sigma^2 Q^{-1} Q Q^{-1}) = \mathcal{N}(0, \sigma^2 Q^{-1})$$

□

t-test statistic:

$$t = \frac{\hat{\beta} - 0}{s.e.(\hat{\beta})}$$

F-test statistic:

$$F = \left(\frac{T - k - 1}{r} \right) \left(\frac{SSR_R - SSR_U}{SSR_U} \right)$$

Since

$$R^2 = \frac{\sum_t (y_t - \overline{y})^2 - \sum_t (y_t - \hat{y}_t)^2}{\sum_t (y_t - \overline{y})^2} = \frac{\sum_t (y_t - \overline{y})^2 - SSR_U}{\sum_t (y_t - \overline{y})^2}$$

we have

$$SSR_U = \sum_t (y_t - \overline{y})^2 - R^2 \sum_t (y_t - \overline{y})^2 = (1 - R^2) \sum_t (y_t - \overline{y})^2$$

yielding

$$F = \left(\frac{T - k - 1}{r} \right) \left(\frac{\sum_t (y_t - \overline{y})^2 - (1 - R^2) \sum_t (y_t - \overline{y})^2}{(1 - R^2) \sum_t (y_t - \overline{y})^2} \right) = \left(\frac{T - k - 1}{r} \right) \left(\frac{R^2}{1 - R^2} \right)$$

Confidence interval for sums of coefficients. (Two coefficient case.) Suppose we want to test $H_0 : \beta_1 + \beta_2 = k$. Let $\delta = \beta_1 + \beta_2 - k$, $\hat{\delta} = \hat{\beta}_1 + \hat{\beta}_2 - k$. Note that under the null hypothesis $\delta = 0$. We can construct a t-statistic

$$t_{\hat{\delta}} = \frac{\hat{\delta} - 0}{\sqrt{\hat{\text{Var}}(\hat{\delta})}} = \frac{\hat{\beta}_1 + \hat{\beta}_2 - k}{\sqrt{\hat{\text{Var}}(\hat{\delta})}}$$

where

$$\hat{\text{Var}}(\hat{\delta}) = \hat{\text{Var}}(\hat{\beta}_1) + \hat{\text{Var}}(\hat{\beta}_2) + 2\hat{\text{Cov}}(\hat{\beta}_1, \hat{\beta}_2)$$

This means that a 95% confidence interval for δ can be constructed in the following way:

$$\hat{\delta} \pm t^* \sqrt{\hat{\text{Var}}(\hat{\delta})}$$

where t^* is the 95% critical value for the t-distribution.

8.4 Chapter 4: Heteroskedasticity

Under heteroskedasticity, the OLS estimator $\hat{\beta} = (X'X)^{-1}X'y$ is unbiased, but the true covariance matrix of $\hat{\beta}$ no longer matches the OLS formula. For instance, suppose we have

$$y_t = \sum_{i=1}^{K} \beta_i x_{ti} + u_t$$

where $\text{Var}(u_t) = \sigma^2 z_t^2$.

$$\hat{\beta} = (X'X)^{-1}X'y = (X'X)^{-1}X'X\beta + (X'X)^{-1}X'u = \beta + (X'X)^{-1}X'u$$

$$\implies \mathbb{E}(\hat{\beta}) = \mathbb{E}[\beta] + (X'X)^{-1}X'\mathbb{E}[u] = \beta$$

since $\mathbb{E}(u)$ is still 0. However,

$$\text{Var}(\hat{\beta}) = \mathbb{E}\big[\big(\hat{\beta} - \mathbb{E}(\hat{\beta})\big)\big(\hat{\beta} - \mathbb{E}(\hat{\beta})\big)'\big] = \mathbb{E}\big[\big(\beta + (X'X)^{-1}X'u - \beta\big)\big(\beta + (X'X)^{-1}X'u - \beta\big)'\big]$$

$$= \mathbb{E}\big[\big((X'X)^{-1}X'u\big)\big((X'X)^{-1}X'u\big)'\big] = \mathbb{E}\big[(X'X)^{-1}X'uu'X\big((X'X)^{-1}\big)'\big]$$

$$= (X'X)^{-1}X'\mathbb{E}[uu' \mid X]X(X'X)^{-1}$$

$$= (X'X)^{-1}X' \begin{bmatrix} \sigma^2 z_1^2 & 0 & 0 & \dots & 0 \\ 0 & \sigma^2 z_2^2 & 0 & \dots & 0 \\ 0 & 0 & \sigma^2 z_3^2 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \sigma^2 z_T^2 \end{bmatrix} X(X'X)^{-1}$$

$$= \sigma^2 (X'X)^{-1}X' \begin{bmatrix} z_1^2 & 0 & 0 & \dots & 0 \\ 0 & z_2^2 & 0 & \dots & 0 \\ 0 & 0 & z_3^2 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & z_T^2 \end{bmatrix} X(X'X)^{-1}$$

which is different from the OLS estimator of the covariance matrix $\sigma^2(X'X)^{-1}$. Therefore the estimate of the variances of $\hat{\beta}$ will be biased if the OLS formulas are used, and the usual t and F tests for $\hat{\beta}$ will be invalid.

8.5 Chapter 5: Autocorrelated disturbances

Generalized least squares model:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u}$$

where

$$\mathbb{E}(\boldsymbol{u} \mid \boldsymbol{X}) = 0 \ \ \forall \ t$$

$$\mathbb{E}(\boldsymbol{u}\boldsymbol{u}' \mid \boldsymbol{X}) = \boldsymbol{\Sigma}$$

where $\boldsymbol{\Sigma}$ is a positive definite matrix.

$$\hat{\beta}_{GLS} = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}y$$

$$\text{Var}(\hat{\beta}_{GLS}) = (X'\Sigma^{-1}X)^{-1}$$

9 Time Series

These notes are based on my notes from *Time Series and Panel Data Econometrics* (1st edition) by M. Hashem Pesaran as well as coursework for Economics 613: Economic and Financial Time Series I at USC.

9.1 Chapter 6: ARDL Models

In an ARDL model, if the error are serially correlated, then the coefficient estimates are biased (even as $T \to \infty$).

9.2 Chapters 12 and 13: Intro to Stochastic Processes and Spectral Analysis

Stationarity conditions: $\{X_t\}$ is **strictly stationary** if the joint distribution functions of $\{X_{t_1}, X_{t_2}, \ldots, X_{t_k}\}$ and $\{X_{t_1+h}, X_{t_2+h}, \ldots, X_{t_k+h}\}$ are identical for all values of $t_1, t_2, \ldots, t_k$ and h and all positive integers k.

X_t is **weakly (or covariance) stationary** if it has a constant mean and variance and its covariance function $\gamma(t_1, t_2)$ depends only on the absolute difference $|t_1 - t_2|$, namely $\gamma(t_1, t_2) = \gamma(|t_1 - t_2|)$.

X_t is said to be **trend stationary** if $y_t = X_t - d_t$ is covariance stationary, where d_t is the perfectly predictable component of X_t.

The process $\{\epsilon_t\}$ is said to be a **white noise process** if it has mean zero, a constant variance, and ϵ_t and ϵ_s are uncorrelated for all $s \neq t$.

Autocovariance generating function: The autocovariance generating function for the general linear stationary process $y_t = \sum_{i=0}^{\infty} a_i \epsilon_{t-i}$ is given by:

$$G(z) = \sigma^2 a(z) a(z^{-1})$$

where $a(z) = \sum_{i=0}^{\infty} a_i z^i$.

Wold's Decomposition (Theorem 42, p. 275, Section 12.5) Any trend-stationary process $\{y_t\}$ can be represented in the form of $y_t = d_t + \sum_{i=0}^{\infty} \alpha_i \epsilon_{t-i}$ where $\alpha_0 = 1$ and $\sum_{i=0}^{\infty} \alpha_i^2 < K < \infty$. The term d_t is a deterministic component, while $\{\epsilon_t\}$ is a serially uncorrelated process: $\epsilon_t = y_t - \mathbb{E}(y_t \mid y_{t-1}, y_{t-2}, \ldots)$.

Stationarity conditions for an ARMA(p, q) process: Consider the ARMA(p, q) process

$$y_t = \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{i=0}^{q} \theta_i \epsilon_{t-i}, \quad \theta_0 = 1$$

The MA part is stationary for any finite q. The AR part is stationary if the roots of the characteristic equation

$$\lambda^t = \sum_{i=1}^{p} \phi_i \lambda^{t-i}$$

lie strictly inside the unit circle. Alternatively, in terms of $z = \lambda^{-1}$, the process is stationary if the roots of

$$1 - \sum_{i=1}^{p} \phi_i z^i = 0$$

lie outside the unit circle. The ARMA process is **invertible** (so that y_t can be solved uniquely in terms of its past values) if all the roots of

$$1 - \sum_{i=1}^{p} \theta_i z^i = 0$$

fall outside the unit circle.

Spectral Density Function: Definition (Equation 13.3):

$$f(\omega) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma(h) e^{ih\omega}, \omega \in (-\pi, \pi)$$

Equation (13.5):

$$f(\omega) = \frac{1}{2\pi} \left[\gamma(0) + 2 \sum_{h=1}^{\infty} \gamma(h) \cos(h\omega) \right], \quad \omega \in [0, \pi]$$

Can also be found using the autocovariance generating function. We have (Equation 13.6, section 13.3.1)

$$f(\omega) = \frac{1}{2\pi} G(e^{i\omega}) = \frac{\sigma^2}{2\pi} a(e^{i\omega}) a(e^{-i\omega})$$

Properties of spectral density function:

(1) $f(\omega)$ always exists and is bounded if $\gamma(h)$ is absolutely summable.

(2) $f(\omega)$ is symmetric.

(3) The spectrum of a stationary process is finite at zero frequency; that is, $f(0) < \infty$.

Linear (time-domain) processes don't have to be stationary, but to write something as a frequency-domain process, it must be stationary.

9.3 Some time series and their properties

9.3.1 White noise process:

$x_t = \epsilon_t$, $\epsilon_t \sim IID(0, \sigma^2)$

- Autocovariances:

$$\gamma(0) = \sigma^2$$

$$\gamma(h) = 0, \quad \forall h \neq 0$$

- Spectral density function:

$$f_x(\omega) = \frac{1}{2\pi} \cdot \sigma^2 = \frac{\sigma^2}{2\pi} \text{ (flat spectrum)}$$

9.3.2 MA(1) process:

$x_t = \epsilon_t + \theta \epsilon_{t-1}$ with $\epsilon_t \sim iid(0, \sigma^2)$, $|\rho| < 1$.

- Autocovariances: By Equation (12.2), the autocovariance function is

$$\text{Cov}(u_t, u_{t-h}) = \gamma(h) = \sigma^2 \sum_{i=0}^{1-|h|} a_i a_{i+|h|} \text{ if } 0 \leq |h| \leq 1$$

$$\implies \mathbb{E}(x_t^2) = \gamma(0) = (1 + \theta^2)\sigma^2$$

$$\mathbb{E}(x_t x_{t-1}) = \gamma(1) = \theta\sigma^2$$

$$\gamma(h) = 0 \;\; \forall \; |h| > 1$$

So the covariance matrix is

$$\begin{pmatrix} \sigma^2(1+\theta^2) & \sigma^2\theta & 0 & 0 & \cdots & 0 \\ \sigma^2\theta & \sigma^2(1+\theta^2) & \sigma^2\theta & 0 & \cdots & 0 \\ 0 & \sigma^2\theta & \sigma^2(1+\theta^2) & \sigma^2\theta & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & \sigma^2\theta & \sigma^2(1+\theta^2) & \sigma^2\theta \\ 0 & 0 & \cdots & 0 & \sigma^2\theta & \sigma^2(1+\theta^2) \end{pmatrix}$$

$$= \sigma^2(1+\theta^2) I_T + \sigma^2 \theta A$$

where A is defined as in section 14.3.2 (p. 304).

- Spectral density function:

$$f(\omega) = \frac{\sigma^2}{2\pi}\left[1 + 2\theta\cos(\omega) + \rho^2\right], \quad \omega \in [0, \pi]$$

9.3.3 MA(∞) process:

This process is covariance stationary.

- Autocovariances:

9.3.4 AR(1) process:

$x_t = \phi x_{t-1} + \epsilon_t$, $|\phi| < 1$, $\epsilon_t \sim IID(0, \sigma^2)$.

- Yule-Walker Equations:

$$\mathbb{E}[x_t x_{t-h}] = \mathbb{E}[\phi x_{t-1} x_{t-h}] + \mathbb{E}[\epsilon x_{t-h}]$$

$$\gamma_h = \phi \gamma_{h-1} + \mathbb{E}[\epsilon x_{t-h}]$$

$$\implies \gamma_0 = \phi \gamma_1 + \sigma^2, \quad \gamma_h = \phi \gamma_{h-1} \; \forall h \geq 1$$

- Autocovariances:

$$\gamma(0) = \frac{\sigma^2}{1 - \phi^2}$$

$$\gamma_h = \frac{\sigma^2 \phi^h}{1 - \rho^2} \quad \forall \; h \geq 1$$

$$\implies \text{Cov}(x) =$$

$$\begin{pmatrix} \sigma^2/(1-\phi^2) & \sigma^2\phi/(1-\phi^2) & \sigma^2\phi^2/(1-\phi^2) & \sigma^2\phi^3/(1-\phi^2) & \cdots & \sigma^2\phi^{T-1}/(1-\phi^) \\ \sigma^2\phi^/(1-\phi) & \sigma^2/(1-\phi^2) & \sigma^2\phi/(1-\phi^2) & \sigma^2\phi^2/(1-\phi^2) & \cdots & \sigma^2\phi^{T-2}/(1-\phi^2) \\ \sigma^2\phi^2/(1-\phi^2) & \sigma^2\phi/(1-\phi^2) & \sigma^2/(1-\phi^2) & \sigma^2\phi/(1-\phi^2) & \cdots & \sigma^2\phi^{T-3}/(1-\phi^2) \\ \vdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ \sigma^2\phi^{T-2}/(1-\phi^2) & \sigma^2\phi^{T-3}/(1-\phi^2) & \cdots & \sigma^2\phi/(1-\phi^2) & \sigma^2/(1-\phi^2) & \sigma^2\phi/(1-\phi^2) \\ \sigma^2\phi^{T-1}/(1-\phi^2) & \sigma^2\phi^{T-2}/(1-\phi^2) & \cdots & \sigma^2\phi^2/(1-\phi^2) & \sigma^2\phi/(1-\phi^2) & \sigma^2/(1-\phi^2) \end{pmatrix}$$

- If stationary, can be written as an infinite MA process with absolutely summable coefficients

$$x_t = \sum_{i=0}^{\infty} \phi^i \epsilon_{t-i} = \left(\frac{1}{1 - \phi L} \right) \epsilon_t$$

- Autocovariance generating function:

$$G(z) = \left(\frac{\sigma^2}{1 - \phi^2} \right) \left(1 + \sum_{h=1}^{\infty} \phi^h (z^h + z^{-h}) \right)$$

- Spectral density function:

$$f(\omega) = \frac{1}{2\pi}\sum_{h=-\infty}^{\infty}\frac{\sigma^2\phi^{|h|}}{(1-\phi^2)}(e^{i\omega})^h = \frac{1}{2\pi}\frac{\sigma^2}{(1-\phi e^{i\omega})(1-\phi e^{-i\omega})} = \frac{1}{2\pi}\frac{\sigma^2}{1-2\phi\cos(\omega)+\phi^2}$$

9.3.5 AR(2) process:

$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \epsilon$, $|\phi_1| < 1$, $|\phi_2| < 1$, $\epsilon_t \sim IID(0, \sigma^2)$.

Can be written as

$$x_t = \frac{1}{1-\phi L}\epsilon_t = \epsilon_t + \phi\epsilon_{t-1} + \phi^2\epsilon_{t-2} + \ldots$$

- Yule-Walker equations:

$$\mathbb{E}[x_t x_{t-h}] = \mathbb{E}[\phi_1 x_{t-1} x_{t-h}] + \mathbb{E}[\phi_2 x_{t-2} x_{t-h}] + \mathbb{E}[\epsilon x_{t-h}]$$

$$\gamma_h = \phi_1\gamma_{h-1} + \phi_2\gamma_{h-2} + \mathbb{E}[\epsilon x_{t-h}]$$

$$\implies \boxed{\gamma_0 = \phi_1\gamma_1 + \phi_2\gamma_2 + \sigma^2, \quad \gamma_1 = \phi_1\gamma_0 + \phi_2\gamma_1, \quad \gamma_2 = \phi_1\gamma_1 + \phi_2\gamma_0}$$

- Autocovariances:

9.3.6 AR(p) process:

$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \ldots + \phi_p x_{t-p} + \epsilon$, $|\phi_i| < 1$, $\epsilon_t \sim IID(0, \sigma^2)$.

- Stationary if the eigenvalues of $\mathbf{\Phi}$ lie inside the unit circle, which is equivalent to all the roots of

  $$\phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \ldots - \phi_p z^p = 0$$

  being strictly larger than unity. Under this condition the AR process has the infinite-order MA representation'

  $$x_t = \sum_{i=0}^{\infty}\alpha_i\epsilon_{t-i}$$

  where $\alpha_i = \phi_1\alpha_{i-1} + \ldots + \phi_p\alpha_{i-p}$.

- Autocovariance generating function:

  $$G(z) = \frac{\sigma^2}{\phi(z)\phi(z^{-1})}$$

9.3.7 ARMA(1, 1) process:

$x_t = \phi x_{t-1} + \epsilon_t + \theta \epsilon_{t-1}$, with $|\phi| < 1$ (implying stationarity), $\mathbb{E}(\epsilon_t^2) = \sigma^2$, $\mathbb{E}(\epsilon_t \epsilon_s) = 0$ for $t \neq s$.

- Yule-Walker Equations:

$$\gamma(0) = \phi\gamma(1) + \sigma^2(1+\theta^2)$$

$$\gamma(1) = \phi\gamma(0) + \sigma^2\phi^2$$

$$\gamma(h) = \phi\gamma(h-1) \;\; \forall\ h \geq 2$$

- Autocovariances:

$$\gamma(0) = \sigma^2\left(1_{\frac{(\phi+\theta)^2}{1-\phi^2}}\right)$$

$$\gamma(1) = \sigma^2\left(\phi + \theta + \frac{(\phi+\theta)^2\phi}{1-\phi^2}\right)$$

$$\gamma(2) = \phi^{h-1}\gamma(1) \;\forall\ h \geq 2$$

- Autocorrelation function:

$$\rho(h) = \begin{cases} 1 & h = 0 \\ \frac{(\phi+\theta)(1+\phi\theta)}{1+2\phi\theta+\theta^2} & h = 1 \\ \phi^{h-1}\rho(1) & h \geq 2 \end{cases}$$

- Autocovariance generating function: the autocovariance function of an ARMA(p, q) process $\phi(L)y_t = \theta(L)\epsilon_t$ is given by

$$f(\omega) = \sigma^2 \frac{\theta(z)\theta(z^{-1})}{\phi(z)\phi(z^{-1})}$$

  Plugging in for the ARMA(1,1) case yields (**double-check**)

$$f(\omega) = \sigma^2 \frac{(1+\theta)^2}{(1-\rho)^2}$$

- Spectral Density Function: the spectral density function of an ARMA(p, q) process $\phi(L)y_t = \theta(L)\epsilon_t$ is given by

$$f(\omega) = \frac{\sigma^2}{2\pi}\frac{\theta(e^{i\omega})\theta(e^{-i\omega})}{\phi(e^{i\omega})\phi(e^{-i\omega})}, \quad \omega \in [0, 2\pi]$$

  Plugging in for the ARMA(1,1) case yields

$$f_x(\omega) = \frac{\sigma^2}{2\pi}\frac{(e^{i\omega} - \theta e^{i\omega})(e^{-i\omega} - \theta e^{-i\omega})}{(e^{i\omega} - \phi e^{i\omega})(e^{-i\omega} - \phi e^{-i\omega})} = \frac{\sigma^2}{2\pi}\frac{1 - 2\theta + \theta^2}{1 - 2\phi + \phi^2}$$

- If $\phi = \theta$, the ARMA(1,1) process becomes a white noise process. We can see this two ways. The ARMA(1, 1) process can be represented in the following way:

$$(1-\phi L)y_t = (1-\theta L)\epsilon_t$$

Therefore $\phi(L) = \theta(L)$ yields $y_t = \epsilon_t$.

We can also see that when $\phi = \theta$, an ARMA(1,1) process is equivalent to a white noise process as follows. Plugging in $\phi = \theta$ to the spectral density function, we have

$$f_x(\omega) = \frac{\sigma^2}{2\pi}\frac{1-2\theta+\theta^2}{1-2\theta+\theta^2} = \frac{\sigma^2}{2\pi}$$

showing that if $\theta = \phi$, the spectral density function is constant and independent of θ and ϕ. We can see that it in fact is a white noise process. Since a white noise process has the following covariances:

$$\gamma(0) = \sigma^2$$

$$\gamma(h) = 0, \quad \forall h \neq 0$$

for a white noise process we have

$$f_x(\omega) = \frac{1}{2\pi}\cdot\sigma^2 = \frac{\sigma^2}{2\pi}$$

9.4 Chapter 14: Estimation of Stationary Time Series Processes

9.4.1 Sufficient conditions for ergodicity of mean. (Book section 14.2.1)

By Chebyshev's Inequality (see section 7.2), $\overline{y}_T$ is a consistent estimator of μ as $T\to\infty$ if $\lim_{T\to\infty}\mathbb{E}(\overline{y}_T) = \mathbb{E}(y_T) = \mu$ and $\lim_{T\to\infty}\text{Var}(\overline{y}_T) = 0$. We have

$$\mathbb{E}(\overline{y}_T) = \frac{1}{T}\mathbb{E}\bigg(\sum_{t=1}^T y_t\bigg) = \frac{1}{T}\sum_{t=1}^T\mathbb{E}(y_t) = \mu$$

$$\text{Var}(\overline{y}_T) = \frac{1}{T^2}\text{Var}\bigg(\sum_{t=1}^T y_t\bigg) = \frac{1}{T^2}\bigg(\sum_{t=1}^T\text{Var}(y_t) + 2\sum_{0\le i<j\le T}\text{Cov}(y_i,y_j)\bigg)$$

$$= \frac{1}{T^2}\bigg(\sum_{t=1}^T\gamma(0) + 2\sum_{0\le i<j\le T}\gamma(j-i)\bigg) = \frac{1}{T^2}\bigg(T\gamma(0) + 2\sum_{h=1}^{T-1}(T-h)\gamma(h)\bigg)$$

$$= \frac{1}{T}\bigg[\gamma(0) + 2\sum_{h=1}^{T-1}\bigg(1-\frac{h}{T}\bigg)\gamma(h)\bigg] = \frac{1}{T^2}\mathbf{1}\text{Var}(\boldsymbol{y})\mathbf{1}'$$

where $\mathbf{1}$ is a vector of ones and

$$\text{Var}(\boldsymbol{y}) = \begin{pmatrix} \gamma(0) & \gamma(1) & \cdots & \gamma(T-2) & \gamma(T-1) \\ \gamma(1) & \gamma(0) & \cdots & \gamma(T-3) & \gamma(T-2) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \gamma(T-2) & \gamma(T-3) & \cdots & \gamma(0) & \gamma(1) \\ \gamma(T-1) & \gamma(T-2) & \cdots & \gamma(1) & \gamma(0) \end{pmatrix}$$

Notice that

$$\left|\gamma(0) + 2\sum_{h=1}^{T-1}\left(1 - \frac{h}{T}\right)\gamma(h)\right| < \left|2\sum_{h=0}^{T-1}\gamma(h)\right| \leq 2\sum_{h=0}^{T-1}|\gamma(h)|$$

Therefore

$$\sum_{h=0}^{T-1}|\gamma(h)| < \infty$$

is a sufficient condition for

$$\lim_{T\to\infty}\text{Var}(\overline{y}_T) = \lim_{T\to\infty}\frac{1}{T}\left[\gamma(0) + 2\sum_{h=1}^{T-1}\left(1 - \frac{h}{T}\right)\gamma(h)\right] = 0$$

9.4.2 Estimation of autocovariances (Book section 14.2.2).

A moment estimator of $\gamma(h) = \mathbb{E}[(y_t - \mu)(y_{t-h} - \mu)]$ is

$$\hat{\gamma}(h) = \frac{1}{T}\sum_{t=h+1}^{T}(y_t - \overline{y}_T)(y_{t-h} - \overline{y}_T)$$

By Chebyshev's Inequality (see section 7.2), $\hat{\gamma}(h)$ is a consistent estimator of $\gamma(h)$ as $T \to \infty$ if $\lim_{T\to\infty}\mathbb{E}(\hat{\gamma}(h)) = \gamma(h)$ and $\lim_{T\to\infty}\text{Var}(\hat{\gamma}(h)) = 0$.

$$\hat{\gamma}(h) = \frac{1}{T}\sum_{t=h+1}^{T}(y_t - \overline{y}_T)(y_{t-h} - \overline{y}_T) = \frac{1}{T}\sum_{t=h+1}^{T}(y_t - \mu + \mu - \overline{y}_T)(y_{t-h} - \mu + \mu - \overline{y}_T)$$

$$= \frac{1}{T}\sum_{t=h+1}^{T}(y_t - \mu)(y_{t-h} - \mu) + (y_t - \mu)(\mu - \overline{y}_T) + (\mu - \overline{y}_T)(y_{t-h} - \mu) + (\mu - \overline{y}_T)^2$$

$$= \frac{1}{T}\sum_{t=h+1}^{T}(y_t - \mu)(y_{t-h} - \mu) + (\mu - \overline{y}_T)\frac{1}{T}\sum_{t=h+1}^{T}(y_t - \mu) + (\mu - \overline{y}_T)\frac{1}{T}\sum_{t=h+1}^{T}(y_{t-h} - \mu) + \frac{1}{T}(T-h)(\mu - \overline{y}_T)^2$$

$$\vdots$$

Because **where does this line come from? on page 300 of book/331 of pdf.**

$$\overline{y}_T = \mu + \mathcal{O}_p(T^{-1/2})$$

and for any fixed h

$$T^{-1/2} \sum_{t=h+1}^{T} (y_t - \mu) = \mathcal{O}_p(1)$$

it follows that

$$(\mu - \overline{y}_T)\frac{1}{T}\sum_{t=h+1}^{T}(y_t - \mu) = \frac{\mu}{T}\sum_{t=h+1}^{T}(y_t - \mu) - \frac{\overline{y}_T}{\sqrt{T}} \cdot \frac{1}{\sqrt{T}}\sum_{t=h+1}^{T}(y_t - \mu) = \mathcal{O}_p(T^{-1})$$

$$(\mu - \overline{y}_T)\frac{1}{T}\sum_{t=h+1}^{T}(y_{t-h} - \mu) = \mathcal{O}_p(T^{-1})$$

$$\frac{1}{T}(T-h)(\mu - \overline{y}_T)^2 = (\mu - \overline{y}_T)^2 - \frac{h}{T}(\mu - \overline{y}_T)^2 = \mathcal{O}_p(T^{-1})$$

$$\implies \hat{\gamma}(h) = \frac{1}{T}\sum_{t=h+1}^{T}(y_t - \mu)(y_{t-h} - \mu) + \mathcal{O}_p(T^{-1})$$

which implies that $\lim_{T\to\infty} \mathbb{E}(\hat{\gamma}(h)) = \gamma(h)$. Also using results in Bartlett (1946) **where? do we need to know how to do this?** we have

$$\lim_{T\to\infty} \text{Var}(\hat{\gamma}_T(h) - \gamma(h)) = 0$$

under the assumption that

$$\lim_{H\to\infty} H^{-1}\sum_{h=1}^{H}\gamma_h^2 \to 0$$

9.4.3 Worked examples

Midterm Problem 2 part (2) (similar to exercise 1 in chapter 14. Suppose $\{y_t\}$ has the following general linear process

$$y_t = \mu + \alpha(L)\epsilon_t, \qquad \epsilon_t \sim i.i.d.\ (0, \sigma^2)$$

where $\alpha(L) = \alpha_0 + \alpha_1 L + \alpha_2 L^2 + \ldots;\ \alpha_0 = 1$. Let

$$\overline{y}_T = \frac{1}{T}\sum_{t=1}^{T} y_t$$

$$\gamma(h) = \mathbb{E}[(y_t - \mu)(y_{t-h} - \mu)]$$

$$\hat{\gamma}(h) = \frac{1}{T}\sum_{t=h+1}^{T}(y_t - \overline{y}_T)(y_{t-h} - \overline{y}_T)$$

Derive the conditions under which

(a) $\overline{y}_T$ is a consistent estimator of μ as $T \to \infty$

(b) For fixed h, $\hat{\gamma}(h)$ is a consistent estimator of $\gamma(h)$ as $T \to \infty$.

Solution.

(a) This is an MA(∞) process. By Chebyshev's Inequality, $\overline{y}_T$ is a consistent estimator of μ as $T \to \infty$ if $\lim_{T\to\infty} \mathbb{E}(\overline{y}_T) = \mathbb{E}(y_T) = \mu$ and $\lim_{T\to\infty} \text{Var}(\overline{y}_T) = 0$. In this case in particular (MA(∞) process), we can write

$$\overline{y}_T = \frac{1}{T}\sum_{t=1}^{T}\big(\mu + \alpha(L)\epsilon_t\big) = \frac{1}{T}\cdot T\mu + \frac{1}{T}\sum_{t=1}^{T}\alpha(L)\epsilon_t = \mu + \frac{1}{T}\sum_{t=1}^{T}\alpha(L)\epsilon_t$$

Then we have

$$\mathbb{E}(\overline{y}_T) = \mu + \frac{1}{T}\mathbb{E}\left(\sum_{t=1}^{T}\alpha(L)\epsilon_t\right) = \mu + \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}(\alpha(L)\epsilon_t) = \mu$$

$$\text{Var}(\overline{y}_T) = 0 + \frac{1}{T^2}\text{Var}\left(\sum_{t=1}^{T}\alpha(L)\epsilon_t\right) = \frac{1}{T^2}\sum_{t=1}^{T}\text{Var}[\alpha(L)\epsilon_t] = \frac{1}{T^2}\sum_{t=1}^{T}\mathbb{E}[\alpha(L)\epsilon_t]^2 = \frac{1}{T}\alpha(1)^2\mathbb{E}[\epsilon_t]^2$$

$$= \frac{\sigma^2}{T}\alpha(1)^2$$

Therefore a sufficient condition for consistency is

$$\lim_{T\to\infty}\frac{\sigma^2}{T}\alpha(1)^2 = 0 \iff \alpha(1)^2 < \infty \impliedby \boxed{\sum_{i=0}^{\infty}\alpha_i = 0}$$

(b) **ask about the derivation** Per the derivation in section 9.4.2, we have that

$$\hat{\gamma}(h) = \frac{1}{T}\sum_{t=h+1}^{T}(y_t - \mu)(y_{t-h} - \mu) + \mathcal{O}_p(T^{-1})$$

For $\hat{\gamma}(h)$ to be consistent, we need

$$\frac{1}{T}\sum_{t=h_1}^{T}(y_t-\mu)(y_{t-h}-\mu) \xrightarrow{p} \gamma(h) \iff \lim_{T\to\infty} \Pr(|\hat{\gamma}(h)-\gamma(h)|<\epsilon)=1, \text{ for every } \epsilon>0$$

First we show that $(y_t-\mu)(y_{t-h}-\mu)$ is a martingale difference process:

$$\mathbb{E}[(y_t-\mu)(y_{t-h}-\mu) \mid F_{t-h}] = (y_{t-h}-\mu)\mathbb{E}[y_t-\mu \mid F_{t-h}] = 0$$

why? which theorem is being used to prove this result? ask We need to show that

$$\mathbb{E}[(y_t-\mu)^2(y_{t-h}-\mu)^2] = \mathbb{E}\left[\left(\sum_{i=0}^{\infty}\alpha_i\epsilon_{t-i}\right)^2\left(\sum_{j=0}^{\infty}\alpha_j\epsilon_{t-h-j}\right)^2\right] < \infty$$

By the Cauchy-Schwarz Inequality, we have

$$\mathbb{E}\left|\left[\left(\sum_{i=0}^{\infty}\alpha_i\epsilon_{t-i}\right)^2\left(\sum_{j=0}^{\infty}\alpha_j\epsilon_{t-h-j}\right)^2\right]\right|^2 \le \mathbb{E}\left[\left(\sum_{i=0}^{\infty}\alpha_i\epsilon_{t-i}\right)^2\right]^2\mathbb{E}\left[\left(\sum_{j=0}^{\infty}\alpha_j\epsilon_{t-h-j}\right)^2\right]^2$$

$$< \infty \iff \mathbb{E}\left[\sum_{i=0}^{\infty}\alpha_i\epsilon_{t-i}\right]^4 < \infty, \quad \mathbb{E}\left[\sum_{j=0}^{\infty}\alpha_j\epsilon_{t-h-j}\right]^4 < \infty$$

These conditions hold if $\mathbb{E}(\epsilon_t^4) < \infty$ and $\sum_{i=0}^{\infty}|\alpha_i| < \infty$. Then $\mathbb{E}[(y_t-\mu)^2(y_{t-h}-\mu)^2] < \infty$ holds and

$$\hat{\gamma} \xrightarrow{p} \gamma(h)$$

Midterm Problem 3 parts (3) and (4) (similar to 14.7 and 14.8 material. Consider the following ARMA(1, 1) model

$$y_t = \phi y_{t-1} + u_t + \theta u_{t-1}, \text{ for } t = -\infty, \dots, -1, 0, 1, \dots$$

where $|\theta| < 1$, $|\phi| < 1$, and u_t is i.i.d. with mean zero and variance σ_u^2, $\mathbb{E}(u_t^4) < \infty$.

(1) Suppose that we have the data $\{y_t : t = 0, 1, \dots, T\}$. Consider the following estimator of ϕ:

$$\hat{\phi}_T = \frac{\sum_{t=2}^{T} y_t y_{t-2}}{\sum_{t=2}^{T} y_{t-1} y_{t-2}}$$

Show that $\hat{\phi}$ is a consistent estimator of ϕ and derive the asymptotic distribution of $\sqrt{T}(\hat{\phi}_T - \phi)$. Comment on the case where $\theta = \phi$.

(2) Suppose that $\sigma_u^2 = 1$ is known. Show that θ can be consistently estimated by

$$\hat{\theta}_T = \frac{1}{T}\sum_{t=1}^{T} y_t y_{t-1} - \frac{\hat{\phi}_T}{T}\sum_{t=1}^{T} y_{t-1}^2$$

Solution.

(1) From the results in Question 2 part 2(b), since $\mathbb{E}(y_t) = \mathbb{E}(y_{t-1}) = \mathbb{E}(y_{t-2}) = 0$, we know that

$$\hat{\phi}_T = \frac{\sum_{t=2}^T y_t y_{t-2}}{\sum_{t=2}^T y_{t-1} y_{t-2}} = \frac{T^{-1}\sum_{t=2}^T y_t y_{t-2}}{T^{-1}\sum_{t=2}^T y_{t-1} y_{t-2}} \xrightarrow{p} \frac{\gamma(2)}{\gamma(1)}$$

By the result from Question 3 part (2), we have $\gamma(h) = \phi\gamma(h-1)$ for $h \geq 2$. Therefore $\gamma(2)/\gamma(1) = \phi$, so $\hat{\phi}_T$ is a consistent estimator for ϕ. To obtain the asymptotic distribution, note that

$$\sqrt{T}(\hat{\phi}_T - \phi) = \sqrt{T}\left(\frac{T^{-1}\sum_{t=2}^T y_t y_{t-2}}{T^{-1}\sum_{t=2}^T y_{t-1} y_{t-2}} - \phi\right)$$

$$= \frac{T^{-1/2}\sum_{t=2}^T (\phi y_{t-1} + u_t + \theta u_{t-1}) y_{t-2}}{T^{-1}\sum_{t=2}^T y_{t-1} y_{t-2}} - \frac{\phi T^{-1/2}\sum_{t=2}^T y_{t-1} y_{t-2}}{T^{-1}\sum_{t=2}^T y_{t-1} y_{t-2}}$$

$$= \frac{T^{-1/2}\sum_{t=2}^T (u_t + \theta u_{t-1}) y_{t-2}}{T^{-1}\sum_{t=2}^T y_{t-1} y_{t-2}}$$

In Question 2 part 2(b), we showed that

$$\frac{1}{T}\sum_{t=h_1}^T (y_t - \mu)(y_{t-h} - \mu) \xrightarrow{p} \gamma(h)$$

Therefore in the denominator, since $\mathbb{E}(y_{t-1}) = \mathbb{E}(y_{t-h})_{=}0$, we have

$$T^{-1}\sum_{t=2}^T y_{t-1} y_{t-2} \xrightarrow{p} \gamma(1)$$

In the numerator,

$$T^{-1/2}\sum_{t=2}^T (u_t + \theta u_{t-1}) y_{t-2} = \frac{1}{\sqrt{T}}\sum_{t=2}^T \left[u_t y_{t-2} + \theta u_{t-1} y_{t-2}\right]$$

$$= \frac{1}{\sqrt{T}}\sum_{t=2}^T u_t y_{t-2} + \frac{1}{\sqrt{T}}\sum_{t=2}^T \theta u_{t-1} y_{t-2} = \frac{1}{\sqrt{T}}\left(\sum_{t=2}^{T-1} u_t y_{t-2} + u_T y_{T-2}\right) + \frac{1}{\sqrt{T}}\sum_{t'=1}^{T-1} \theta u_{t'} y_{t'-1}$$

$$= \frac{1}{\sqrt{T}}\left(\sum_{t=2}^{T-1} u_t y_{t-2} + u_T y_{T-2}\right) + \frac{1}{\sqrt{T}}\left(\theta u_1 y_0 + \sum_{t=2}^{T-1} \theta u_t y_{t-1}\right) = \frac{1}{\sqrt{T}}\left(\sum_{t=2}^{T-1} u_t (y_{t-2} + \theta y_{t-1}) + \theta u_1 y_0 + u_T y_{T-2}\right)$$

Since $\mathbb{E}(u_t(y_{t-2} + \theta y_{t-1}) \mid F_{t-1}) = 0$. Further, $T^{-1/2}(\theta u_1 y_0 + u_{T-1} y_{T-2}) = o_p(1)$. Then by the Central Limit Theorem in martingale difference processes (see section 73):

Theorem 28 (Central limit theorem for martingale difference sequences). Let $\{x_t\}$ be a martingale difference sequence with respect to the information set Ω_t. Let $\overline{\sigma}_T^2 = \text{Var}(\sqrt{T}\overline{x}_T) = T^{-1}\sum_{t=1}^T \sigma_t^2$. If $\mathbb{E}(|x_t|^r) < K < \infty$, $r > 2$ and for all t, and

$$\frac{1}{T}\sum_{t=1}^T x_t^2 - \overline{\sigma}_t^2 \xrightarrow{p} 0$$

then

$$\sqrt{T} \cdot \frac{\overline{x}_T}{\overline{\sigma}_T} \xrightarrow{d} \mathcal{N}(0,1)$$

we have

$$\sqrt{T} \cdot \frac{\overline{x}_T}{T^{-1/2}\sqrt{\sum_{t=1}^T \sigma_t^2}} \xrightarrow{d} \mathcal{N}(0,1)$$

$$\vdots$$

$$\frac{1}{\sigma^2}\frac{\gamma(1)^2}{(1+\theta)^2)\gamma(0)+2\theta\gamma(1)}\sqrt{T}(\hat{\phi}_T - \phi) \xrightarrow{d} \mathcal{N}(0,1)$$

$$\iff \sqrt{T}(\hat{\phi}_T - \phi) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2\frac{(1+\theta)^2)\gamma(0)+2\theta\gamma(1)}{\gamma(1)^2}\right)$$

(2) From the results of Question 2 part 2(b), where we showed that

$$\frac{1}{T}\sum_{t=h_1}^T (y_t - \mu)(y_{t-h} - \mu) \xrightarrow{p} \gamma(h)$$

(and since $\mathbb{E}(y_{t-1}) = \mathbb{E}(y_{t-h})_=0$,)

$$T^{-1}\sum_{t=2}^T y_t y_{t-1} \xrightarrow{p} \gamma(1), \quad T^{-1}\sum_{t=2}^T y_{t-1}^2 \xrightarrow{p} \gamma(0)$$

and by the law of large numbers (see section 7.2) **(why?)**, we have

$$\hat{\theta}_T = \frac{1}{T}\sum_{t=1}^T y_t y_{t-1} - \frac{\hat{\phi}_T}{T}\sum_{t=1}^T y_{t-1}^2 \xrightarrow{p} \gamma(1) - \phi\gamma(0) = \phi\gamma(0) + \theta\sigma^2 - \phi\gamma(0) = \theta$$

9.5 Chapter 17: Introduction to Forecasting

9.5.1 17.7: Iterated and direct multi-step AR methods

Suppose y_t follows the AR(1) model:

$$y_t = a + \phi y_{t-1} + \epsilon_t, \quad |\phi| < 1, \epsilon_t \sim iid(0, \sigma_\epsilon^2) \tag{1}$$

$$\iff y_t = \frac{a}{1-\phi} + \sum_{i=0}^{\infty} \phi^i \epsilon_{t-i}$$

$$\iff y_t = a\left(\frac{1-\phi^h}{1-\phi}\right) + \phi^h y_{t-h} + \sum_{j=0}^{h-1} \phi^j \epsilon_{t-j} \tag{2}$$

We have two methods for forecasting y_{t+h} $h > 1$ steps ahead.

(1) **Iterated method:** In this method, we first calculate the OLS estimates of $\hat{a}_T$ and $\hat{\phi}_T$ in Equation (1) using all available data Ω_T. Then we use the form of Equation (2):

$$\hat{y}^*_{T+h|T} = \hat{a}_T\left(\frac{1-\hat{\phi}_T^h}{1-\hat{\phi}_T}\right) + \hat{\phi}_T^h y_T$$

(2) **Direct method:** We directly calculate OLS estimates of the parameters in Equation (2) using all available data Ω_T:

$$\tilde{y}^*_{T+h|T} = \tilde{a}_{h,T} + \tilde{\phi}_{h,T} y_T$$

Proposition 45. Suppose data is generated by Equation (1). If $u_t = \sum_{i=0}^{\infty} \phi^i \epsilon_{t-i}$ and $v_t = \sum_{j=0}^{h-1} \phi^j \epsilon_{t-j}$ are symmetrically distributed around zero and have finite second moments, and if $\mathbb{E}(\hat{\phi}_T$ and $\mathbb{E}(\tilde{\phi}_{h,T})$ exist, then for any finite T and h we have

$$\mathbb{E}(\hat{y}^*_{T+h|T} - y_{T+h}) = \mathbb{E}(\tilde{y}^*_{T+h|T} - y_{T+h}) = 0$$

10 Abstract Algebra

These are my notes from reading *Elementary Abstract Algebra* by W. Edwin Clark, available for free download on his website: http://shell.cas.usf.edu/~wclark/#ELEMENTARY_ABSTRACT_ALGEBRA

10.1 Chapter 1: Binary Operations

Definition 1.1 *A* **binary operation** $*$ *on a set* S *is a function from* $S \times S$ *to* S. *If* $(a, b) \in S \times S$ *then we write* $a*b$ *to indicate the image of the element* (a, b) *under the function* $*$.

The following lemma explains in more detail exactly what this definition means.

Lemma 1.1 *A binary operation* $*$ *on a set* S *is a rule for combining two elements of* S *to produce a third element of* S. *This rule must satisfy the following conditions:*

(a) $a \in S$ *and* $b \in S \Longrightarrow a * b \in S$. [$S$ is closed under $*$.]

(b) *For all* a, b, c, d *in* S
$a = c$ *and* $b = d \Longrightarrow a * b = c * d$. [Substitution is permissible.]

(c) *For all* a, b, c, d *in* S
$a = b \Longrightarrow a * c = b * c$.

(d) *For all* a, b, c, d *in* S
$c = d \Longrightarrow a * c = a * d$.

Definition: A **function** f from the set A to the set B is a rule which assigns to each element $a \in A$ an element $f(a) \in B$ in such a way that the following condition holds for all $x, y \in A$:

$$x = y \implies f(x) = f(y)$$

To indicate that f is a function from A to B we write $f : A \to B$. The set A is called the **domain** of f and the set B is called the **codomain** of f.

A function $f : A \to B$ is said to be **one-to-one** or **injective** if the following condition holds for all $x, y \in A$:

$$f(x) = f(y) \implies x = y$$

A function $f : A \to B$ is said to be **onto** or **surjective** if the following condition holds:

$$\forall\ b \in B\ \exists\ a \in A \mid f(a) = b$$

A function $f : A \to B$ is said to be **bijective** if it is both one-to-one and onto. Then f is sometimes said to be a **bijection** or a **one-to-one correspondence** between A and B.

15. Let S, T, and U be nonempty sets, and let $f: S \to T$ and $g: T \to U$ be functions such that the function $g \circ f: S \to U$ is one-to-one (injective). Which of the following must be true?

(A) f is one-to-one.

(B) f is onto.

(C) g is one-to-one.

(D) g is onto.

(E) $g \circ f$ is onto.

Solution 15. (A) For a composition of functions, if the first function isn't one-to-one, there's no way the composite is. It's worth mentioning here that the opposite is true for onto: the second function had better be onto.

Let S be a set. The **power set** $\mathcal{P}(S)$ of S is the set of all subsets of S (including S itself).

Definition 1.2 *Assume that $*$ is a binary operation on the set S.*

1. *We say that $*$ is* **associative** *if*

$$x * (y * z) = (x * y) * z \quad \text{for all } x, y, z \in S.$$

2. *We say that an element e in S is an* **identity** *with respect to $*$ if*

$$x * e = x \text{ and } e * x = x \quad \text{for all } x \text{ in } S.$$

3. *Let $e \in S$ be an identity with respect to $*$. Given $x \in S$ we say that an element $y \in S$ is an* **inverse** *of x if both*

$$x * y = e \text{ and } y * x = e.$$

4. *We say that $*$ is* **commutative** *if*

$$x * y = y * x \quad \text{for all } x, y \in S.$$

5. *We say that an element a of S is* **idempotent** *with respect to $*$ if*

$$a * a = a.$$

6. *We say that an element z of S is a* **zero** *with respect to $*$ if*

$$z * x = z \text{ and } x * z = z \quad \text{for all } x \in S.$$

For each integer $n \geq 2$ define the set

$$\mathbb{Z}_n = \{0, 1, 2, \dots, n-1\}$$

For all $a, b \in \mathbb{Z}_n$ let

$$a + b = \text{remainder when the ordinary sum of a and b is divided by n}$$

and

$$a \cdot b = \text{remainder when the ordinary product of a and b is divided by n.}$$

These binary operations are referred to as **addition modulo** n and **multiplication modulo** n. The integer n in $\mathbb{Z}_n$ is called the **modulus**. The plural of modulus is **moduli**.

Let K denote any one of the following: $\mathbb{Z}, \mathbb{Q}, \mathbb{R}, \mathbb{Z}_n$.

$$M_n(K)$$

is the set of all $n \times n$ matrices containing elements of K.

$$GL(n, K)$$

is the set of all matrices in $M_n(K)$ with non-zero determinant. $(GL(n,k), \cdot)$ is called the **general linear group of degree n over K**. It is non-abelian.

$$SL(n, K) = \{A \in GL(n, K) \mid \det(A) = 1\}$$

$SL(n, K)$ is called the **Special Linear Group of degree** n **over** K.

10.2 Chapter 2: Groups

Definition A **group** is an ordered pair $(G, *)$ where G is a set and $*$ is a binary operation on G satisfying the following properties:

1. **The binary operation is associative on G:** $\forall\ x, y, z \in G$,
$$x * (y * z) = (x * y) * z$$
2. **The binary operation contains a (unique) identity in G:** $\exists\ e \in G \mid \forall\ x \in G$
$$e * x = x,\ x * e = x$$
3. **Every element in G has a (unique) inverse on $*$ in G:** $\forall\ x \in G\ \exists\ y \in G \mid$
$$x * y = e, y * x = e$$

A group $(G, *)$ is said to be **abelian** if $\forall\ x, y \in G,\ x * y = y * x$. A group is said to be **non-abelian** if it is not abelian.

Theorem 2.2 *Let $(G, *)$ be a group with identity e. Then the following hold for all elements a, b, c, d in G:*

1. *If $a * c = a * b$, then $c = b$.* [Left cancellation law for groups.]
2. *If $c * a = b * a$, then $c = b$.* [Right cancellation law for groups.]
3. *Given a and b in G there is a* unique *element x in G such that $a*x = b$.*
4. *Given a and b in G there is a* unique *element x in G such that $x*a = b$.*
5. *If $a * b = e$ then $a = b^{-1}$ and $b = a^{-1}$.* [Characterization of the inverse of an element.]
6. *If $a * b = a$ for just one a, then $b = e$.*
7. *If $b * a = a$ for just one a, then $b = e$.*
8. *If $a * a = a$, then $a = e$.* [The only idempotent in a group is the identity.]
9. $(a^{-1})^{-1} = a$.
10. $(a * b)^{-1} = b^{-1} * a^{-1}$.

10.3 Chapter 3: The Symmetric Groups

If n is a positive integer,

$$[n] = \{1, 2, \dots, n\}$$

A **permutation** of $[n]$ is a one-to-one, onto function from $[n]$ to $[n]$, and

$$S_n$$

is the set of all permutations of $[n]$.

The identity of S_n is the so-called **identity function**

$$\iota : [n] \to [n]$$

which is defined by the rule

$$\iota(x) = x, \quad \forall\, x \in [n]$$

The inverse of an element $\sigma \in S_n$: Suppose $\sigma \in S_n$. Since σ is by definition one-to-one and onto, the rule

$$\sigma^{-1}(y) = x \iff \sigma(x) = y$$

defines a function $\sigma^{-1} : [n] \to [n]$. This function σ^{-1} is also one-to-one and onto and satisfies

$$\sigma\sigma^{-1} = \iota \text{ and } \sigma^{-1}\sigma = \iota$$

so it is the inverse of σ in the group sense also.

Since the binary operation of composition on S_n is associative $[(\gamma\beta)\alpha = \gamma(\beta\alpha)]$, S_n under the binary operation of composition is a group (it is associative, it has an inverse, and it has an identity).

Definition 3.2 *Let $i_1, i_2, \ldots, i_k$ be a list of k distinct elements from $[n]$. Define a permuation σ in S_n as follows:*

$$\begin{array}{rcl} \sigma(i_1) & = & i_2 \\ \sigma(i_2) & = & i_3 \\ \sigma(i_3) & = & i_4 \\ \vdots & \vdots & \vdots \\ \sigma(i_{k-1}) & = & i_k \\ \sigma(i_k) & = & i_1 \end{array}$$

and if $x \notin \{i_1, i_2, \ldots, i_k\}$ then

$$\sigma(x) = x$$

Such a permutation is called a **cycle** *or a* k-**cycle** *and is denoted by*

$$(i_1\ i_2\ \cdots\ i_k).$$

If $k = 1$ then the cycle $\sigma = (i_1)$ is just the identity function, i.e., $\sigma = \iota$.

Two cycles $(i_1\ i_2\ \ldots\ i_k)$ and $(j_1\ j_2\ \ldots\ j_l)$ are said to be **disjoint** if the sets $\{i_1, i_2, \ldots, i_k\}$ and $\{j_1, j_2, \ldots, j_l\}$ are disjoint.

So for example, the cycles (1 2 3) and (4 5 8) are disjoint, but the cycles (1 2 3) and (4 2 8) are not disjoint.

If σ and τ are disjoint cycles, then $\sigma\tau = \tau\sigma$.

Theorem 3.4 *Every element $\sigma \in S_n$, $n \geq 2$, can be written as a product*

$$\sigma = \sigma_1\sigma_2\cdots\sigma_m \tag{3.1}$$

where $\sigma_1, \sigma_2, \ldots, \sigma_m$ are pairwise disjoint cycles, that is, for $i \neq j$, σ_i and σ_j are disjoint. If all 1-cycles of σ are included, the factors are unique except for the order. ∎

The factorization (3.1) is called the **disjoint cycle decomposition of** σ.

An element of S_n is called a **transposition** if and only if it is a 2-cycle.

Every element of S_n can be written as a product of transpositions. The factors of such a product are not unique. However, if $\sigma \in S_n$ can be written as a product of k transpositions and if the same σ can also be written as a product of l transpositions, then k and l have the same parity.

A permutation is **even** if it is a product of an even number of transpositions and **odd** if it is a product of an odd number of transpositions. We define the function sign : $S_n \to \{1, -1\}$ by

$$\text{sign}(\sigma) = \begin{cases} 1 & \text{if } \sigma \text{ is even} \\ -1 & \text{if } \sigma \text{ is odd} \end{cases}$$

If $n = 1$ then there are no transpositions. In this case, to be complete we define the identity permutation ι to be even.

If σ is a k-cycle, then $\text{sign}(\sigma) = 1$ if k is odd and $\text{sign}(\sigma) = -1$ if k is even.

Remark. Let $A = [a_{ij}]$ be an $n \times n$ matrix. The determinant of A may be defined by the sum

$$\det(A) = \sum_{\sigma \in S_n} \text{sign}(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{n\sigma(n)}.$$

For example, if $n = 2$ we have only two permutations ι and $(1\ 2)$. Since $\text{sign}(\iota) = 1$ and $\text{sign}((1\ 2)) = -1$ we obtain

$$\det(A) = a_{11}a_{22} - a_{12}a_{21}.$$

Definition: If $(G, *)$ is a group, the number of elements in G is called the **order** of G. We use $|G|$ to denote the order of G. Note that $|G|$ may be finite or infinite.

Let

$$A_n$$

be the set of all even permutations in the group S_n. A_n is called the **alternating group of degree** n.

10.4 Chapter 4: Subgroups

Definition: Let G be a group. A **subgroup** of G is a subset H of G which satisfies the following three conditions:

1. $e \in H$
2. $a, b \in H \implies ab \in H$
3. $a \in H \implies a^{-1} \in H$

If H is a subgroup of G, we write $H \leq G$. The subgroups $\{e\}$ and G are said to be **trivial** subgroups of G.

Every finite subgroup may be thought of as a subgroup of one of the groups S_n.

Let A_n be the set of all even permutations in the group S_n. A_n is then a subgroup of S_n. A_n is called the **alternating group of degree** n.

Let a be an element of the group G. If $\exists\ n \in \mathbb{N} \mid a^n = e$ we say that a has **finite order** and we define

$$\mathrm{o}(a) = \min\{n \in \mathbb{N} \mid a^n = e\}$$

If $a^n \neq e\ \forall\ n \in \mathbb{N}$ we say that a has **infinite order** and we define

$$\mathrm{o}(a) = \infty$$

In either case we call $\mathrm{o}(a)$ the **order** of a. Note carefully the difference between the order of a group and the order of an element of a group. Note also that $a = e \iff \mathrm{o}(a) = 1$. So every element of a group other than e has order $n \geq 2$ or ∞.

Let a be an element of group G. Define

$$\langle a \rangle = \{a^i : i \in \mathbb{Z}\}$$

We call $\langle a \rangle$ the **subgroup of** G **generated by** a. Note that $e = a^0$ and a^{-1} are in $\langle a \rangle$.

Theorem. For each $a \in G$, $\langle a \rangle$ is a subgroup of G. $\langle a \rangle$ contains a and is the smallest subgroup of G containing a.

Proof of second statement. If H is any subgroup of G containing a, $\langle a \rangle \subseteq H$ since H is closed under taking products and inverses. That is, every subgroup of G containing a also contains $\langle a \rangle$. This implies that $\langle a \rangle$ is the smallest subgroup of G containing a.

Theorem. Let G be a group and let $a \in G$. If $\mathrm{o}(a) = 1$, then $\langle a \rangle = \{e\}$. If $\mathrm{o}(a) = n$ where $n \geq 2$, then

$$\langle a \rangle = \{e, a, a^2, \ldots, a^{n-1}\}$$

and the elements $e, a, a^2, \ldots, a^{n-1}$ are distinct; that is,

$$\mathrm{o}(a) = |\langle a \rangle|$$

Proof Assume that $\mathrm{o}(a) = n$. The case $n = 1$ is left to the reader. Suppose $n \geq 2$. We must prove two things.

1. If $i \in \mathbb{Z}$ then $a^i \in \{e, a, a^2, \ldots, a^{n-1}\}$.
2. The elements $e, a, a^2, \ldots, a^{n-1}$ are distinct.

To establish 1 we note that if i is any integer we can write it in the form $i = nq + r$ where $r \in \{0, 1, \ldots, n-1\}$. Here q is the quotient and r is the remainder when i is divided by n. Now using Theorem 2.4 we have

$$a^i = a^{nq+r} = a^{nq}a^r = (a^n)^q a^r = e^q a^r = ea^r = a^r.$$

This proves 1. To prove 2, assume that $a^i = a^j$ where $0 \leq i < j \leq n-1$. It follows that

$$a^{j-i} = a^{j+(-i)} = a^j a^{-i} = a^i a^{-i} = a^0 = e.$$

But $j - i$ is a positive integer less than n, so $a^{j-i} = e$ contradicts the fact that $\mathrm{o}(a) = n$. So the assumption that $a^i = a^j$ where $0 \leq i < j \leq n-1$ is false. This implies that 2 holds. It follows that $\langle a \rangle$ contains exactly n elements, that is, $\mathrm{o}(a) = |\langle a \rangle|$.

Theorem. If G is a finite group, then every element of G has finite order.

49. What is the largest order of an element in the group of permutations of 5 objects?

(A) 5 (B) 6 (C) 12 (D) 15 (E) 120

Solution 49. (B) The greatest order is given by the product of a 2-cycle and a 3-cycle acting on disjoint elements. That gives order 6.

10.5 Chapter 5: The Group of Units of $\mathbb{Z}_n$

Let $n \geq 2$. An element $a \in \mathbb{Z}_n$ is said to be a **unit** if $\exists\, b \in \mathbb{Z}_n \mid ab = 1$ (where the product is multiplication modulo n).

The set of all units in $\mathbb{Z}_n$ is denoted by

$$U_n$$

and is a group under multiplication modulo n called the **group of units of $\mathbb{Z}_n$**.

Theorem. For $n \geq 2$, $U_n = \{a \in \mathbb{Z}_n : \gcd(a, n) = 1\}$

Theorem. p is a prime $\implies \exists\, a \in U_p \mid U_p = \langle a \rangle$

Theorem. If $n \geq 2$ then U_n contains an element a satisfying $U_n = \langle a \rangle$ if and only if a has one of the following forms: $2,\ 4,\ p^k$, or $2p^k$ where p is an odd prime and $k \in \mathbb{N}$.

10.6 Chapter 6: Direct Products of Groups

If $G_1, G_2, \ldots, G_n$ is a list of n groups we make the Cartesian product $G_1 \times G_2 \times \cdots \times G_n$ into a group by defining the binary operation

$$(a_1, a_2, \ldots, a_n) \cdot (b_1, b_2, \ldots, b_n) = (a_1 \cdot b_1, a_2 \cdot b_2, \ldots, a_n \cdot b_n)$$

Here for each $i \in \{1, 2, \ldots, n\}$ the product $a_i \cdot b_i$ is the product of a_i and b_i in the group G_i. We call this group the **direct product** of the groups $G_1, G_2, \ldots, G_n$.

The direct product contains an identity and an inverse, and is associative (since it is composed of groups which must themselves be associative), so it is a group per below:

Theorem. If $G_1, G_2, \ldots, G_n$ is a list of n groups, the direct product $G = G_1 \times G_2 \times \cdots \times G_n$ as defined above is a group. Moreover, if for each i, e_i is the identity of G_i, then $e_1, e_2, \ldots, e_n$ is the identity of G, and if

$$\boldsymbol{a} = (a_1, a_2, \ldots, a_n) \in G$$

then the inverse of $\boldsymbol{a}$ is given by

$$\boldsymbol{a}^{-1} = (a_1^{-1}, a_2^{-1}, \ldots, a_n^{-1})$$

where a_i^{-1} is the inverse of a_i in the group G_i.

10.7 Chapter 7: Isomorphism of Groups

Let $G = \{g_1, g_2, \ldots, g_n\}$. Let $\text{o}(g_i) = k_i$ for $i = 1, 2, \ldots, n$. We say that the sequence $(k_1, k_2, \ldots, k_m)$ is the **order sequence** of the group G. To make the sequence unique we assume the elements are ordered so that $k_1 \leq k_2 \leq \ldots \leq k_n$.

Let $(G, *)$ and $(H, \bullet)$ be groups. A function $f : G \to H$ is said to be a **homomorphism** from G to H if

$$f(a * b) = f(a) \bullet f(b)$$

for all $a, b \in G$. If in addition f is one-to-one and onto, f is said to be an **isomorphism** from G to H.

We say that G and H are **isomorphic** if and only if there is an isomorphism from G to H. We write $G \cong H$ to indicate that G is isomorphic to H.

Isomorphism is an equivalence relation: If G, H, and K are groups then

1. $G \cong G$
2. If $G \cong H$ then $H \cong G$, and
3. If $G \cong H$ and $H \cong K$, then $G \cong K$.

Theorem. Let $(G, *)$ and $(H, \bullet)$ be groups and let $f : G \to H$ be a homomorphism. Let e_G denote the identity of G, and let e_H denote the identity of H. Then

1. $f(e_G) = e_H$

   Proof: Let $x_G \in G$ and let $f(x_G) = x_H \in H$. Then
   $x_H = f(x_G) = f(e_G * x_G) = f(e_G) \bullet f(x_G) = f(e_G) \bullet x_H = e_H \bullet x_H$.

2. $f(a^{-1}) = f(a)^{-1}$

   Proof: $f(a)^{-1} \bullet f(a) = e_H = f(e_G) = f(a^{-1} * a) = f(a^{-1}) \bullet f(a)$

3. $f(a^n) = f(a)^n \ \forall \ n \in \mathbb{Z}$

   Proof by induction.

Theorem. Let $(G, *)$ and $(H, \bullet)$ be groups and let $f : G \to H$ be an isomorphism. Then $\text{o}(a) = \text{o}(f(a)) \ \forall \ a \in G$. It follows that G and H have the same number of elements of each possible order.

Theorem. If G and H are isomorphic groups, and G is abelian, then so is H.

Proof: Let $a_G, b_G \in G$ and let $f(a_G) = a_H \in H, f(b_G) = b_H \in H$.
$a_H \bullet b_H = f(a_G) \bullet f(b_G) = f(a_G * b_G) = f(b_G * a_G) = f(b_G) \bullet f(a_G) = b_H \bullet a_H$.

A group G is **cyclic** if there is an element $a \in G \mid \langle a \rangle = G$. If $\langle a \rangle = G$ then we say that a is a **generator** for G.

Theorem. If G and H are isomorphic groups and G is cyclic then H is cyclic.

Theorem. Let a be an element of group G.

1. $\mathrm{o}(a) = \infty \implies \langle a \rangle \cong \mathbb{Z}$.
2. $\mathrm{o}(a) = n \in \mathbb{N} \implies \langle a \rangle \cong \mathbb{Z}_n$

Cayley's Theorem. If G is a finite group of order n, then there is a subgroup H of S_n such that $G \cong H$.

66. Let $\mathbb{Z}_{17}$ be the ring of integers modulo 17, and let $\mathbb{Z}_{17}{}^{\times}$ be the group of units of $\mathbb{Z}_{17}$ under multiplication. Which of the following are generators of $\mathbb{Z}_{17}{}^{\times}$?

    I. 5

    II. 8

    III. 16

    (A) None (B) I only (C) II only (D) III only (E) I, II, and III

Solution 66. (B) We need to pick elements of order 16 in $\mathbb{Z}/17^{\times}$. It is easy to rule out $16 \equiv -1$, since -1 has order 2. We see that $5^2 = 25 \equiv 8$, so there's no way that 8 can be a generator. We just need to verify that the order of 5 is more than 8, so we can check 5^8:

$$5^4 = 8^2 = 64 \equiv -4, \quad 5^8 = (-4)^2 = 16 \neq 1.$$

That makes 5 a generator.

10.8 Chapter 8: Cosets and Lagrange's Theorem

Let G be a group and let H be subgroup of G. For each element a of G we define

$$aH = \{ah \mid h \in H\}$$

We call aH the **coset of H in G generated by** a.

Let $a, b \in G$. Then

1. $a \in aH$ (since H must contain an identity; specifically, the identity of G)
2. $|aH| = |H|$ (since ah is unique)
3. $aH \cap bH \neq \emptyset \implies aH = bH$

Lagrange's Theorem. If G is a finite group and $H \leq G$ then $|H|$ divides $|G|$.

Any group of prime order is cyclic; therefore, there is only one such group up to isomorphism.

Exercise 3. Use Lagrange's theorem to prove that any group of prime order is cyclic.

Proof. Let G be a group whose order is a prime p. Since $p > 1$, there is an element $a \in G$ such that $a \neq e$. The group $\langle a \rangle$ generated by a is a subgroup of G. By Lagrange's theorem, the order of $\langle a \rangle$ divides $|G|$. But the only divisors of $|G| = p$ are 1 and p. Since $a \neq e$ we have $|\langle a \rangle| > 1$, so $|\langle a \rangle| = p$. Hence $\langle a \rangle = G$ and G is cyclic. $\square$

We say that there are k **isomorphism classes of groups of order** n if there are k groups $G_1, G_2, \ldots, G_k$ such that

1. if $i \neq j$ then G_i and G_j are not isomorphic, and
2. Every group of order n is isomorphic to G_i for some $i \in \{1, 2, \ldots, k\}$.

This is sometimes expressed by saying that "there are k groups of order n up to isomorphism" or that "there are k non-isomorphic groups of order n."

12. For which integers n such that $3 \leq n \leq 11$ is there only one group of order n (up to isomorphism) ?

(A) For no such integer n

(B) For 3, 5, 7, and 11 only

(C) For 3, 5, 7, 9, and 11 only

(D) For 4, 6, 8, and 10 only

(E) For all such integers n

Solution 12. (B) Any group of prime order is necessarily cyclic, and hence there is only one up to isomorphism. This limits are choices to (B), (C), and (E). But there are two groups of order 9 (at least): $\mathbb{Z}/3 \times \mathbb{Z}/3$ and $\mathbb{Z}/9$. This makes (B) our only option.

In more advanced courses in algebra, it is shown that the number of isomorphism classes of groups of order n for $n \leq 17$ is given by the following table:

Order :	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Number :	1	1	1	2	1	2	1	5	2	2	1	5	1	2	1	14	1

This table means, for example, that one may find 14 groups of order 16 such that every group of order 16 is isomorphic to one and only one of these 14 groups.

There is only one isomorphism class of groups of order n if n is prime. But there are some non-primes that have this property; for example, 15.

The Fundamental Theorem of Finite Abelian Groups. If G is a finite abelian group of order at least 2, then

$$G \cong \mathbb{Z}_{p_1^{n_1}} \times \mathbb{Z}_{p_2^{n_2}} \times \cdots \times \mathbb{Z}_{p_s^{n_s}}$$

where for each i, p_i is a prime and n_i is a positive integer. Moreover, the prime powers $p_i^{n_i}$ are unique except for the order of the factors.

If the group G in the above theorem has order n then

$$n = p_1^{n_1} p_2^{n_2} \cdots p_s^{n_2}$$

So the p_i may be obtained from the prime factorization of the order of the group G. These primes are not necessarily distinct, so we cannot say what the n_i are. However, we can find all possible choices for the n_i. For example, if G is an abelian group of order $72 = 3^2 \cdot 2^3$ then G is isomorphic to one and only one of the following groups. Note that each corresponds to a way of factoring 72 as a product of prime powers.

$$\begin{array}{ll}
\mathbb{Z}_9 \times \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2 & 72 = 9 \cdot 2 \cdot 2 \cdot 2 \\
\mathbb{Z}_9 \times \mathbb{Z}_4 \times \mathbb{Z}_2 & 72 = 9 \cdot 4 \cdot 2 \\
\mathbb{Z}_9 \times \mathbb{Z}_8 & 72 = 9 \cdot 8 \\
\mathbb{Z}_3 \times \mathbb{Z}_3 \times \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2 & 72 = 3 \cdot 3 \cdot 2 \cdot 2 \cdot 2 \\
\mathbb{Z}_3 \times \mathbb{Z}_3 \times \mathbb{Z}_4 \times \mathbb{Z}_2 & 72 = 3 \cdot 3 \cdot 4 \cdot 2 \\
\mathbb{Z}_3 \times \mathbb{Z}_3 \times \mathbb{Z}_8 & 72 = 3 \cdot 3 \cdot 8
\end{array}$$

Thus there are exactly 6 non-isomorphic abelian groups of order 72.

Corollary. For $n \geq 2$, the number of isomorphism classes of abelian groups of order n is equal to the number of ways to factor n as a product of prime powers (where the order of the factors does not count).

10.9 Chapter 9: Introduction to Ring Theory

Definition 9.1 *A **ring** is an ordered triple $(R, +, \cdot)$ where R is a set and $+$ and $\cdot$ are binary operations on R satisfying the following properties:*

A1 $a + (b + c) = (a + b) + c$ *for all* a, b, c *in* R.

A2 $a + b = b + a$ *for all* a, b *in* R.

A3 *There is an element* $0 \in R$ *satisfying* $a + 0 = a$ *for all* a *in* R.

A4 *For every* $a \in R$ *there is an element* $b \in R$ *such that* $a + b = 0$.

M1 $a \cdot (b \cdot c) = (a \cdot b) \cdot c$ *for all* a, b, c *in* R.

D1 $a \cdot (b + c) = a \cdot b + a \cdot c$ *for all* a, b, c *in* R.

D2 $(b + c) \cdot a = b \cdot a + c \cdot a$ *for all* a, b, c *in* R.

Terminology If $(R, +, \cdot)$ is a ring, the binary operation $+$ is called *addition* and the binary operation $\cdot$ is called *multiplication*. *In the future we will usually write ab instead of* $a \cdot b$. The element 0 mentioned in A3 is called the **zero** of the ring. Note that we have not assumed that 0 behaves like a *zero*, that is, we have not assumed that $0 \cdot a = a \cdot 0 = 0$ for all $a \in R$. What A3 says is that 0 is an identity with respect to addition. Note that *negative* (as the opposite of *positive*) has no meaning for most rings. We do not assume that multiplication is commutative and we have not assumed that there is an identity for multiplication, much less that elements have inverses with respect to multiplication.

23. Let $(\mathbb{Z}_{10}, +, \cdot)$ be the ring of integers modulo 10, and let S be the subset of $\mathbb{Z}_{10}$ represented by $\{0, 2, 4, 6, 8\}$. Which of the following statements is FALSE?

    (A) $(S, +, \cdot)$ is closed under addition modulo 10.

    (B) $(S, +, \cdot)$ is closed under multiplication modulo 10.

    (C) $(S, +, \cdot)$ has an identity under addition modulo 10.

    (D) $(S, +, \cdot)$ has no identity under multiplication modulo 10.

    (E) $(S, +, \cdot)$ is commutative under addition modulo 10.

Solution 23. (D) Examining the choices, we see $S \subset \mathbb{Z}/10$ is a subgroup of an abelian group. Therefore it still have an additive identity and the operation is commutative. It is also closed under addition and multiplication. While S does not contain the multiplicative identity of $\mathbb{Z}/10$, it does have a multiplicative identity. $6 \in S$ is such an identity, as

$$6x = (5+1)x = 5x + x.$$

Since $x \in S$ are all even, $5x = 0$, so $6x = x$.

50. Let R be a ring and let U and V be (two-sided) ideals of R. Which of the following must also be ideals of R ?

I. $U + V = \{u + v : u \in U \text{ and } v \in V\}$

II. $U \cdot V = \{uv : u \in U \text{ and } v \in V\}$

III. $U \cap V$

(A) II only (B) III only (C) I and II only (D) I and III only (E) I, II, and III

Solution 50. (D) The sum of the ideals is still an ideal: it is clearly closed under addition (using commutativity of addition), and still under left and right multiplication due to the distributive property. The intersection of ideals is still an ideal, which is not too hard to work out. The product of ideals, however, need not be closed under addition. Consider, for example, $R = \mathbb{Z}[X]$, $U = (2, X)$, and $V = (3, X)$ (the ideals generated by two elements). Then we know that $-2X \in U \cdot V$ and $3X \in U \cdot V$, and hence we should expect $3X - 2X = X \in U \cdot V$. However, there is no way to get X as the product of an element of U and an element of V.

18. Let V be the real vector space of all real 2×3 matrices, and let W be the real vector space of all real 4×1 column vectors. If T is a linear transformation from V onto W, what is the dimension of the subspace $\{\mathbf{v} \in V : T(\mathbf{v}) = \mathbf{0}\}$?

(A) 2 (B) 3 (C) 4 (D) 5 (E) 6

Solution 18. (A) We see that $\dim V = 6$ and $\dim W = 4$. Since $\dim \operatorname{im} T = \dim W = 4$, we must have $\dim \ker T = 6 - 4 = 2$.

11 Miscellaneous

6. Which of the following circles has the greatest number of points of intersection with the parabola $x^2 = y + 4$?

(A) $x^2 + y^2 = 1$

(B) $x^2 + y^2 = 2$

(C) $x^2 + y^2 = 9$

(D) $x^2 + y^2 = 16$

(E) $x^2 + y^2 = 25$

Solution 6. (C) We can try to do this algebraically, but non-algebraically is simpler. Graphing $y = x^2 - 4$ shows that the graph crosses the x-axis at ± 2. Therefore a circle of radius 1 or $\sqrt{2}$ will not intersect the parabola at all. A circle of radius 3 will intersect four times – twice above and twice below the x-axis. A circle of radius 4 will only intersect at one point below the x-axis (and twice above), and a circle of radius 5 will only intersect at the two points above.

19. If z is a complex variable and $\overline{z}$ denotes the complex conjugate of z, what is $\lim_{z \to 0} \frac{(\overline{z})^2}{z^2}$?

(A) 0 (B) 1 (C) i (D) ∞ (E) The limit does not exist.

Solution 19. (E) Let us represent $z = a + bi$. Then our limit becomes

$$\lim_{(a,b) \to 0} \frac{(a - bi)^2}{(a + bi)^2} = \lim_{(a,b) \to 0} \frac{a^2 - b^2 - 2abi}{a^2 - b^2 + 2abi}.$$

If we let $a = 0$ (for instance), it is easy to see that the limit is equal to 1. However, if we let $a = b$, then our limit becomes

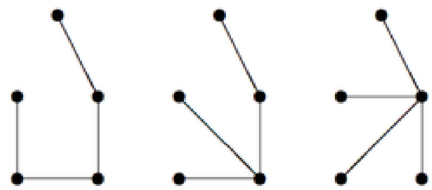
$$\lim_{a \to 0} \frac{-2a^2 i}{2a^2 i} = -1.$$

Therefore the limit does not exist.

29. A tree is a connected graph with no cycles. How many nonisomorphic trees with 5 vertices exist?

(A) 1 (B) 2 (C) 3 (D) 4 (E) 5

Solution 29. (C) It's probably easiest to draw this out for yourself. The maximum degree of any vertex is 2, 3, or 4. If there is a vertex of degree 4, then our tree looks like a star. If the maximum degree of any vertex is 2, then we have a straight line. In the middle case, we obtain a 3-pointed star to which we attach one more vertex – the choice of branch yields isomorphic graphs. See Figure 1.



38. The maximum number of acute angles in a convex 10-gon in the Euclidean plane is

(A) 1 (B) 2 (C) 3 (D) 4 (E) 5

Solution 38. (C) The total angle measure of a 10-gon is $180 \cdot 8 = 1440°$. If the polygon is to be convex, all angles must be less than 180°. If we have 5 acute angles, then the remaining 5 angles would have to make up for $> 1440 - 5 \cdot 90 = 990$ degrees. This is impossible to do and remain convex. If we have 4 acute angles, the remaining 6 angles need to make up for $> 1440 - 4 \cdot 90 = 1080$ degrees. This is our edge case, so the answer must be 3 acute angles.

45. How many positive numbers x satisfy the equation $\cos(97x) = x$?

(A) 1 (B) 15 (C) 31 (D) 49 (E) 96

Solution 45. (C) Certainly our solutions are concentrated in $[0, 1]$. We know that every $2\pi/97$ units in x, we get another period of $\cos(97x)$, and each period must meet $y = x$ twice. Therefore there are

$$\frac{1}{2\pi/97} = \frac{97}{2\pi} \approx \frac{97}{6.3} \approx 15$$

periods in $[0, 1]$ and about 30 meetings. There's only one answer in that range, so we'll stick with it.