**차 례**

머리말 6

제1장 우연사건과 확률 7

1.1 우연사건과 그 연산 7

1.1.1 우연현상 7

1.1.2 표본공간 7

1.1.3 우연사건 7

1.1.4 우연량 9

1.1.5 사건들사이의 관계 10

1.1.6 사건들사이의 연산 11

1.1.7 사건령역 12

1.2 확률의 정의와 그 결정방법 16

1.2.1 확률의 공리적정의 17

1.2.2 배렬과 조합에 대한 공식 17

1.2.3 확률결정의 잦음률방법 19

1.2.4 확률결정의 고전적방법 21

1.2.5 확률결정의 기하방법 28

1.2.6 확률을 결정하는 주관적방법 31

1.3 확률의 성질 34

1.3.1 확률의 가법성 34

1.3.2 확률의 단조성 35

1.3.3 확률의 더하기공식 36

1.3.4 확률의 련속성 38

1.4 조건부확률 42

1.4.1 조건부확률의 정의 42

1.4.2 곱하기공식 44

1.4.3 완전확률공식 46

1.4.4 베이스공식 48

1.5 독립성 53

1.5.1 두 사건의 독립성 53

1.5.2 여러 사건의 호상독립성 54

1.5.3 실험의 독립성 58

제2장 우연량과 그 분포 61

2.1 우연량과 그 분포 61

2.1.1 우연량의 개념 61

2.1.2 우연량의 분포함수 62

2.1.3 리산우연량의 확률분포렬 65

2.1.4 련속우연량의 확률밀도함수 68

2.2 우연량의 수학적기대값 76

2.2.1 수학적기대값의 개념 76

2.2.2 수학적기대값의 정의 77

2.2.3 수학적기대값의 성질 80

2.3 우연량의 분산과 표준편차 85

2.3.1 분산과 표준편차의 정의 85

2.3.2 분산의 성질 88

2.3.3 체븨쉐브부등식 88

2.4 대표적인 리산분포 91

2.4.1 2항분포 91

2.4.2 뽜쏭분포 94

2.4.3 초기하분포 99

2.5대표적인 련속분포 104

2.5.1 정규분포 104

2.5.2 평등분포 109

2.5.3 지수분포 110

2.5.4 감마분포 112

2.5.5 베타분포 114

2.6 우연량함수의 분포 120

2.6.1 리산우연량함수의 분포 120

2.6.2 련속우연량함수의 분포 120

2.7분포의 기타 특징수 127

2.7.1 차모멘트 127



2.7.2 변이곁수 128

2.7.3 분위수 128

2.7.4 중위수 130

2.7.5 비대칭도 131

2.7.6 뾰족도 133

제3장 다차원우연량과 그 분포 137

3.1 다차원우연량과 그 동시분포 137

3.1.1 다차원우연량 137

3.1.2 동시분포함수 137

3.1.3 동시분포렬 140

3.1.4 동시밀도함수 141

3.1.5 대표적인 다차원분포 142

3.2 주변분포와 우연량의 독립성 149

3.2.1 주변분포함수 149

3.2.2 주변분포렬 150

3.2.3 주변밀도함수 151

3.2.4 우연량들사이의 독립성 153

3.3 다차원우연량함수의 분포 158

3.3.1 다차원리산우연량함수의 분포 158

3.3.2 최대값과 최소값의 분포 161

3.3.3 련속인 경우의 합성적공식 163

3.3.4 변수변환법 166

3.4 다차원우연량의 특징수 169

3.4.1 다차원우연량함수의 수학적기대값 170

3.4.2 수학적기대값과 분산의 연산에 관한 성질 171

3.4.3 공분산 174

3.4.4 상관곁수 177

3.4.5 우연벡토르의 수학적기대벡토르와 공분산행렬 182

3.5 조건부분포와 조건부기대값 188

3.5.1 조건부분포 188

3.5.2 조건부수학적기대값 193

제4장 큰수의 법칙과 중심극한정리 200

4.1 우연량렬의 두가지 수렴성 200

4.1.1 확률수렴 200

4.1.2 분포수렴, 약수렴 202

4.2 특징함수 206

4.2.1 특징함수의 정의 206

4.2.2 특징함수의 성질 208

4.2.3 특징함수에 의한 분포함수의 유일결정 211

4.3 큰수의 법칙 216

4.3.1 베르누이의 큰수법칙 216

4.3.2 대표적인 몇가지 큰수의 법칙 219

4.4 중심극한정리 224

4.4.1 독립우연량들의 합 224

4.4.2 독립동일분포하는 경우의 중심극한정리 226

4.4.3 2항분포의 정규근사 228

4.4.4 독립이지만 분포가 서로 다른 경우의 중심극한정리 231

제5장 통계량과 그분포 237

5.1모집단과 표본 237

5.1.1 모집단과 개체 237

5.1.2 표본 239

5.2 표본자료의 정리와 표시 241

5.2.1 경험적분포함수 241

5.2.2 빈도수표와 잦음률표 242

5.2.3 표본자료의 도형표현 243

5.3 통계량과 그 분포 246

5.3.1 통계량과 표본화분포 246

5.3.2 표본평균과 그 표본분포 247

5.3.3 표본분산과 표준편차 250

5.3.4 표본모멘트와 그 함수 252

5.3.5 순서통계량과 그 분포 254

5.3.6 표본분위수와 표본중위수 258

5.3.7 5개수 요약과 상자선도 259

5.4 3대표본분포 265

5.4.1 분포 266

5.4.2 분포 268

5.4.3 분포 270

5.5 충분통계량 275

5.5.1 충분성의 개념 275

5.5.2 인자분해정리 278

제6장 파라메터추정 283

6.1 점추정의 개념과 불편성 283

6.1.1 점추정과 불편성 283

6.1.2 유효성 286

6.2 모멘트추정과 일치성 288

6.2.1 치환원리와 모멘트추정 288

6.2.2 확률함수가 기지일 때 미지파라메터의 모멘트추정 289

6.2.3 일치성 290

6.3 최대우도추정과 EM알고리듬 293

6.3.1 최대우도추정 293

6.3.2 EM알고리듬 296

6.3.3 점근정규성 298

6.4 최소분산불편추정 302

6.4.1 평균두제곱오차 303

6.4.2 평등최소분산불편추정 304

6.4.3 충분성준칙 306

6.4.4 크라메르-라오부등식 307

6.5 베이스추정 311

6.5.1 통계적추정의 기초 311

6.5.2 베이스공식의 밀도함수형식 312

6.5.3 베이스추정 313

6.5.4 공액사전분포 315

6.6 구간추정 317

6.6.1 구간추정의 개념 317

6.6.2 중심축량법 320

6.6.3 한개 정규모집단파라메터에 대한 믿음구간 321

6.6.4 큰표본의 믿음구간 324

6.6.5 표본량 확정 325

6.6.6 두 정규모집단에 대한 믿음구간 327

제7장 가설검정 333

7.1 가설검정의 기본사상과 개념 333

7.1.1 가설검정문제 333

7.1.2 가설검정의 기본단계 334

7.1.3 검정의 값 339

7.2 정규모집단파라메터의 가설검정 341

7.2.1 한개 정규모집단평균의 검정 341

7.2.2 가설검정과 믿음구간사이의 관계 347

7.2.3 두 정규모집단평균차의 검정 348

7.2.4 쌍자료검정 350

7.3 다른 분포파라메터에 대한 가설검정 359

7.3.1 지수분포파라메터에 대한 가설검정 359

7.3.2 비률 의 검정 360

7.3.3 큰표본검정 361

7.4 우도비검정과 분포정합검정 364

7.4.1 우도비검정의 원리 364

7.4.2 분류자료의 정합도검정 366

7.4.3 분포의 정합도검정 369

7.4.4 분할표의 독립성검정 371

7.5 정규성검정 376

7.5.1 정규확률그라프 376

7.5.2 검정 379

7.5.3 EP검정 382

7.6 비파라메터검정 384

7.6.1 순회검정 384

7.6.2 부호검정 387

7.6.3 순위합검정 390

제8장 분산분석과 회귀분석 396

8.1 분산분석 396

8.1.1 문제설정 396

8.1.2 단일인자분산분석의 통계적모형 396

8.1.3 두제곱합분해 398

8.1.4 검정방법 400

8.1.5 파라메터추정 403

8.1.6 중복수가 같지 않은 경우 404

8.2 다중비교 408

8.2.1 수준평균차의 믿음구간 408

8.2.2 다중비교문제 409

8.2.3 중복수가 같은 경우의 T방법 410

8.2.4 중복수가 서로 다른 경우의 S방법 411

8.3 분산동일성검정 413

8.3.1 하틀리검정 413

8.3.2 바틀리트검정 415

8.3.3 수정된 바틀리트검정 418

8.4 1차원선형회귀 419

8.4.1 변수들사이의 두가지 관계 419

8.4.2 1차원선형회귀모형 420

8.4.3 회귀곁수의 최소두제곱추정 422

8.4.4 회귀방정식의 유의성검정 424

8.4.5 추정과 예측 430

8.5 1차원비선형회귀 437

8.5.1 가능한 함수형식의 확정 437

8.5.2 파라메터추정 439

8.5.3 곡선회귀방정식의 비교 441

부록 444

련습문제답안 465

참고문헌 481

# 머리말

확률 및 수리통계학은 모든 대학들의 수학, 통계학과의 기초과목이다. 이 책의 목적은 풍부한 배경과 독창적인 사고, 흥미로운 결론으로 독자들과 학생들이 관심을 가지고 확률 및 수리통계의 기본개념들과 방법, 리론을 배우고 숙달할수 있도록 하는것이다. 개념과 결론에 대한 설명에 중점을 두어 이 학문에 대한 기초지식이 없는 학생이 쉽게 배우고 교원들이 가르치기 쉽도록 내용을 구성하였다.

이 책의 내용은 모두 8개장으로서 처음 4개장은 확률론, 마지막 4개장은 수리통계학이다. 우리는 우연량의 정의를 두 단계로 나누어 주었는데 1장에서는 직관적인 정의를 주어 사건을 표시하였으며 2장에서는 좀 더 엄격한 정의를 제시하였다. 이를 통해 우연량에 대한 보다 구체적이고 완전한 개념을 가질수 있다. 우연량에 대하여 분포의 개념을 더 강조하였다. 또한 확률의 정의에 대하여 공리계를 리용하였으며 잦음률, 고전적확률, 기하확률, 주관적확률을 확률을 결정하는 4가지 방법으로 취하였다. 수리통계부분에서는 가능한껏 자료로부터 출발하여 문제를 제기하고 연구하도록 하였으며 모집단, 표본분포, 검정의 거절령역 등의 개념을 구체적으로 설명하였다. 서술통계학의 기본내용들과 베이스통계학의 기초적인 내용들을 주어 학생들이 비교적 전면적으로 통계학을 인식할수 있도록 하였다. 분위수, 검정의 값, 점근분포 등에 대하여 가능한껏 구체적으로 설명하였다. 이 과정에 100여개의 그림을 리용하여 내용들을 쉽게 리해할수 있도록 하였다.

확률 및 수리통계학의 입문도서로서 학생들이 이 학문에 접하자마자 수학의 심오한 리론세계로 유도하는것이 아니라 현실로부터 확률과 통계의 다양한 배경을 파악한 상태에서 여러가지 개념과 정리를 주고 엄밀한 증명, 그에 대한 직관적인 해석도 따라세우면서 전통적인 수학교과서들과 다른 리해방식을 추구하였다.

이 책에 제시된 실례는 거의 250개에 달하며 그중 많은것이 사람들의 사회, 경제, 생활, 생산관리에 더 접근하고 최근 현실을 반영하고 있으며 일상적인 교육과 연구에서 수집되였다.

이 책에 절별로 준 련습문제를 통하여 해당 개념과 결론들을 더 집중적으로 리해할수 있고 지식을 확장할수 있으며 그 과정에 학생들은 흥미를 더 강하게 느끼고 자신감을 가질수 있다. 련습문제의 수를 많이 늘였는데 책에는 600개 이상의 문제가 있다. 그중에서 절반은 대부분의 학생들이 기본지식만 습득하면 풀수 있는 기본문제이고 나머지 부분은 사고를 많이 해야 풀수 있는 문제들이다.

이 책을 교과서로 리용하여 확률 및 수리통계학을 두 학기에 60시간씩 2학기에 걸쳐 강의할수 있다. 한 학기에 강의를 끝마치려는 경우 일부 내용을 선택하여 강의를 할수 있다. 이때 확률론부분에서 1장, 2장의 대부분 내용, 수학적기대값과 분산의 연산성질, 베르누이의 큰수법칙, 중심극한정리 등을 선택할수 있으며 통계학부분에서 5장, 6장, 7장의 대부분을 선택하고 충분통계량, 최소분산불편추정, 두표본가설검정 등은 생략할수 있다.

# 제1장 우연사건과 확률

확률론과 수리통계학의 연구대상은 우연현상이며 확률론은 우연현상에 대한 모형(즉 확률분포)과 그 성질을 연구한다. 상세한 내용은 앞의 4개장을 참고할수 있다. 수리통계는 우연현상에 대한 자료수집과 처리, 통계적추론을 연구한다. 자세한 내용은 뒤의 4개장을 참고할수 있다. 우연현상으로부터 시작하여 점차적으로 이 내용들을 설명한다.

## 1.1 우연사건과 그 연산

### 1.1.1 우연현상

동전던지기나 주사위던지기와 같이 일정한 조건하에서 항상 같은 결과가 나타나지 않는 현상을 **우연현상**이라고 한다. 우연현상은 두가지 특성을 가진다.

(1) 결과가 하나뿐이 아니다.

(2) 어떤 결과가 나올지 사전에 알수 없다.

단 하나의 결과만 나타나는 현상을 **확정성현상**이라고 한다. 례를 들면 해가 동쪽에서 솟고 물이 낮은 곳으로 흐르며 다른 부호의 전하들이 서로 끌어당기고 한 주머니속에 완전히 같은 흰색공이 10개 있을 때 임의로 꺼낸것은 반드시 흰색공이 된다는것 등을 들수 있다.

실례 1.1.1 우연현상의 실례.

(1) 동전을 던질 때 앞면이 우로 향할수도 있고 뒤면이 우로 향할수도 있다.

(2) 주사위를 던졌을 때 나타나는 점수.

(3) 하루동안 어느한 상점에 들어간 손님수.

(4) 어떤 텔레비죤의 수명.

(5) 물리적량(길이, 직경 등)을 측정한 오차.

우연적인 현상은 어디에나 있다.

동일한 조건에서 반복될수 있는 우연현상에 대한 관측, 기록, 실험을 **우연실험**이라고 한다. 례를 들면 어떤 축구경기의 승패는 반복될수 없으며 일부 경제현상(실업, 경제성장속도 등)도 반복될수 없다. 확률론과 수리통계는 주로 다량의 중복될수 있는 우연현상을 연구하지만 중복되지 않는 우연현상의 연구도 중시한다.

### 1.1.2 표본공간

우연현상의 가능한 모든 기본결과들의 모임을 **표본공간**이라고 하며 로 표시한다. 여기서 는 기본결과를 나타내는데 **표본점**이라고도 부른다. 표본점은 앞으로의 표본발취에서 가장 기본적인 단위이다. 우연현상을 인식하려면 우선 그것의 표본공간을 렬거해야 한다.

실례 1.1.2 실례 1.1.1에서 보여준 우연현상의 표본공간을 렬거하시오.

(1) 동전던지기의 표본공간은 인데 여기서 은 앞면이 우로 향했다는것을, 은 반대면이 우로 향했다는것을 나타낸다.

(2) 주사위던지기의 표본공간은 인데 여기서 .은 점수 가 출현하는것을 나타낸다. 간단히 이 표본공간을 으로 표시할수도 있다.

(3) 하루에 어느한 상점에 들어온 손님수의 표본공간은 이다. 여기서 '0'은 하루동안 아무도 상점을 찾지 않았다는 뜻이고 ''은 하루동안 10만명이 상점을 찾았다는 뜻이다. 비록 이 두가지 상황이 아주 적게 발생하지만 발생하지 않는다고 말할수 없으며 지어는 하루동안 상점에 들어가는 최대 인원수까지도 정확히 말할수 없으므로 표본공간은 실제 상황에도 부합되고 수학적처리에도 편리하도록 부가 아닌 옹근수로 표시한다.

(4) TV수명의 표본공간은 이며 여기서 는 TV가 동작하기 시작해서부터 처음 고장이 발생할 때까지의 시간이다.

(5) 측정오차의 표본공간은 , 여기서 는 측정값 와 참값 사이의 차이다. 즉 .

다음과 같은 점에 주의를 돌려야 한다.

(1) 표본공간의 원소는 수일수도 있고 아닐수도 있다.

(2) 우연현상의 표본공간에는 적어도 두개의 표본점이 있다. 만일 확정적인 현상을 함께 고려한다면 하나의 표본점을 포함하는 표본공간이 확정적현상에 대응한다.

(3) 표본공간을 표본점의 개수로 구분하면 유한 및 무한표본공간의 두 종류로 나눌수 있다. 례를 들어 우의 표본공간 의 표본점개수는 유한개이고 의 표본점개수는 무한개이다. 그러나 에서는 표본점의 개수를 셀수 있지만 와 에서는 표본점의 개수가 셀수 없는 무한개이다. 흔히 표본점의 개수가 유한하거나 셀수 있는 경우를 한가지로 분류하는데 이것을 **리산표본공간**이라고 하고 표본점의 개수가 셀수 없는 무한개인 경우를 다른 부류로 분류하여 **련속표본공간**이라고 한다. 이 두 종류의 표본공간은 본질적인 차이가 있다.

### 1.1.3 우연사건

우연현상의 일부 표본점들로 구성된 모임을 **우연사건**, 간단히 **사건**이라고 부르고 대문자 로 표시한다. 만일 주사위를 던질 때 "홀수점수가 출현한다"는 하나의 사건으로서 {1, 3, 5}로 표시되는데 그것은 표본공간 {1, 2, 3, 4, 5, 6}의 하나의 부분모임이다. 이러한 사건의 정의에서 다음의 몇가지를 주의하여야 한다.

(1) 임의의 사건 는 해당한 표본공간의 부분모임이다. 확률론에서는 흔히 그림 1.1.1에 보여준 **벤도식**(Venn graph)과 같이 하나의 직4각형으로 표본공간 를 표시하고 그안의 원이나 다른 기하도형으로 사건 를 표시한다.

(2) 부분모임 에서 어느 표본점이 출현하면 사건 가 발생한다고 한다. 또는 사건 가 발생할 때에만 의 어떤 표본점이 출현한다고 말한다.

(3) 사건은 모임으로 표시할수도 있고 명백하게 언어로 서술할수도 있다.

(4) 표본공간 의 매개 원소로 구성되는 부분모임을 **기본사건**이라고 한다. 표본공간 의 가장 큰 부분모임(자체)을 **필연사건**이라고 부른다. 표본공간 의 최소부분모임(빈모임 )을 **불가능사건**이라고 부른다.









그림 1.1.1 사건 의 벤도식

실례1.1.3 주사위던지기의 표본공간은 ={1, 2, …, 6}이다.

사건 ="점수 1이 출현한다"는 의 하나의 표본점 "1"로 구성된다.

사건 ="짝수점수가 출현한다"는 의 세개의 표본점 "2, 4, 6"으로 구성된다.

사건 ="7보다 작은 점수가 출현한다"는 의 모든 표본점 "1, 2, 3, 4, 5, 6"으로 구성되며 필연사건으로 된다.

사건 ="6보다 큰 점수가 출현한다"에 대하여 의 아무런 표본점도 에 속하지 않으며 따라서 는 빈모임이고 불가능사건 이다.

### 1.1.4 우연량

우연현상의 결과를 표시하는 변수를 **우연량**이라고 하는데 흔히 대문자 로 표시한다. 많은 사건들은 모두 우연량으로 표시할수 있는데 이때 우연량의 의미를 명확히 밝혀야 한다. 그리고 우연량의 의미는 사람들의 요구에 따라 설정한다. 이것을 아래에서 몇가지 례를 통하여 설명하겠다.

실례1.1.4 많은 우연현상의 결과는 수이다. 이러한 수를 어떤 변수의 값으로 본다면 우연량을 얻을수 있다. 실례로 주사위던지기에서는 1, 2, 3, 4, 5, 6이 출현할수 있다. 만일 ="주사위를 던져 출현하는 점수"로 설정하면 1, 2, 3, 4, 5, 6은 우연량 의 가능한 값이다. 이때

• 사건 "3점이 출현한다"는 "=3"으로 표시한다.

• 사건 "출현점수가 3점을 초과한다"는 ">3"으로 표시한다.

• ""은 필연사건 이다.

• ""은 불가능사건 이다.

이 우연현상에서 "주사위를 한번 던질 때 점수 6점이 출현하는 회수"라고 설정한다면 는 0이나 1의 값만을 취하는 우연량이며 와 다른 새로운 우연량이다. 이때

• ""은 사건 "6점이 출현하지 않는다"을 나타낸다.

• ""은 사건 "6점이 출현한다"를 나타낸다.

• ""은 필연사건 이다.

• ""는 불가능사건 이다.

이와 같이 한가지 우연현상에 대해서도 어떻게 설정하는가에 따라 서로 다른 우연량을 얻을수 있으며 그 설정은 필요에 따라 진행될수 있다.

실례1.1.5 일부 우연현상의 결과는 수가 아니지만 필요에 따라 의미있는 우연량을 설정할수 있다. 례하면 하나의 제품을 검사할 때 합격품과 불합격품의 두가지 결과가 나올수 있다. 만일 불합격품에 주의를 둔다면 ="한개 제품을 검사한 결과 얻은 불합격품의 수"로 설정할수 있는데 이때 는 0과 1만을 취하는 우연량이며 "=0"은 사건 "합격품이 출현한다"를 나타내고 "=1"은 사건 "불합격품이 출현한다"를 나타낸다.

만일 10개의 제품을 검사할 경우 불합격품수 는 하나의 우연량으로서 0, 1, 2, …, 10의 11개 값을 취할수 있다. 그리고

• 사건 "불합격품수가 1개보다 많지 않다"는 ""로 나타낸다.

• ""은 사건 "모두 합격품이다"를 나타낸다.

• ""은 사건 "모두 불합격품이다"를 나타낸다.

• ""은 불가능사건 를 나타낸다.

• ""은 필연사건 이다.

이로부터 알수 있다싶이 우연량은 사람들이 연구와 수요에 근거하여 설정한것으로서 이것을 같기부호, 안같기부호를 리용하여 어떤 실수들과 련결시키면 많은 사건을 표시할수 있다. 이런 표시방법은 형식이 간단하고 의미가 명확하며 리용하기 편리하다. 앞으로 많은 사건들은 모두 우연량으로 표시하게 되는데 여기서 중요한것은 우연량의 설정에 대하여 미리 설명해야 한다는것이다.

### 1.1.5 사건들사이의 관계

이제부터는 항상 한가지 표본공간 (한가지 우연현상)에 대하여 론의를 진행한다고 가정한다. 사건들사이의 관계는 모임사이의 관계와 마찬가지로 주로 다음과 같은 몇가지가 있다.

1. 포함관계

만일 에 속하는 표본점이 반드시 에 속해야 한다면 는 에 포함된다, 또는 가 를 포한한다고 하고(그림 1.1.2 참고)  또는 로 표시한다. 확률론의 언어로 "사건 가 발생하면 사건 가 반드시 발생한다"와 같이 말할수 있다.

례를 들어 주사위던지기에서 사건 ="4점이 출현한다"는 필연적으로 사건 ="짝수점수가 출현한다"의 발생을 초래한다. 따라서 . 또한 텔레비죤의 수명 가 10, 000시간을 초과한다는 사건 과 가 20, 000시간을 초과한다는 사건 에 대하여 이라고 말할수 있다(그림 1.1.3을 참고). 임의의 사건 에 대하여 이다.













0

10000

20000

그림1.1.2 

그림1.1.3 

2. 동등관계

만일 사건 와 사건 에 대하여 에 속하는 표본점이 에 속하고 또 에 속하는 표본점은 에 속하면 즉 이면 사건 와 는 동등하다고 하고 로 표시한다.

모임론의 관점에서 볼 때 두 사건이 동등하다는것은 이 두 사건이 같은 모임이라는것을 의미한다. 다음의 례는 서로 다르게 묘사된 사건이 때로는 동등한 사건일수도 있다는것을 보여준다.

실례 1.1.6 주사위를 두번 던질 때 를 사건 "주사위를 두번 던진 점수의 합은 홀수이다"로, 를 사건 "주사위를 두번 던진 점수가 하나는 홀수이고 다른 하나는 짝수이다"라고 하면 라는것을 쉽게 알수 있다.

실례 1.1.7 주머니에는 검은색공 a개와 흰색공 b개(a, b>0)가 있고 공을 다 꺼낼 때까지 다시 넣지 않고 하나씩 꺼낸다. 사건 가 "마지막으로 꺼낸 여러 개의 공은 모두 검은색공이다"이고 사건 가 "마지막으로 꺼낸 1개의 공이 검은색공이다"고 하면 얼핏 보기에 인것 같지만 공을 전부 꺼낼 때까지 생각하면 가 발생하면 반드시 가 발생하게 된다. 즉 . 반대로 사건 에서 지적된 "여러 개"가 최소 1개, 또는 2개일수 있으며 많아서 a개도 될수 있다는것을 생각하면 사건 의 발생이 반드시 사건 의 발생을 초래한다. 즉 이며 따라서 로 된다.

3. 배반관계

만일 와 가 그림 1.1.4와 같이 동일한 표본점을 가지고있지 않다면 사건 와 는 서로 배반이라고 말한다. 확률론의 언어로 말한다면 사건 와 가 배반이라는것은 사건 와 가 동시에 발생할수 없다는것이다.





그림 1.1.4 사건 와 의 배반관계

례를 들어TV수명에 대한 시험에서 두 사건 "수명이 1만시간미만이다"와 "수명이 5만시간이상이다"는 동시에 발생할수 없는 사건들로서 서로 배반사건들이다.

### 1.1.6 사건들사이의 연산

사건의 연산과 모임의 연산이 서로 대응하는데 합, 사귐, 차, 나머지의 4가지의 연산이 있다.

1. 사건 와 의 합

로 표기한다. 이것은 사건 와 의 모든 표본점(동일한것은 한번만 계산된다)으로 구성되는 새로운 사건을 의미한다 (그림 1.1.5). 확률론적언어로 "사건 와 가운데서 적어도 하나는 발생한다."라고 말한다.

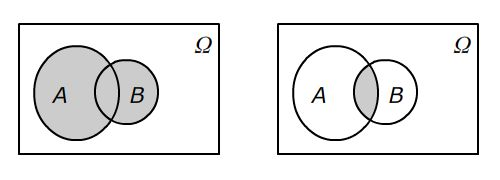


그림 1.1.6 사건 와 의 사귐

그림 1.1.6 사건 와 의 합

한개의 주사위를 던지는 시험에서 사건 ="홀수점수가 출현한다"={1, 3, 5}, 사건 ="점수가 3을 넘지 않는다"={1, 2, 3}이면 와 의 합은 ={1, 2, 3, 5}이다.

2. 사건 와 의 사귐

 또는 로 표기한다. 이것은 사건 와 의 공통적인 표본점으로 구성되는 새로운 사건을 의미한다(그림 1.1.6). 확률론적언어로 "사건 와 가 동시에 발생한다"라고 말한다.

한개의 주사위를 던지는 시험에서 사건 ="홀수점수가 출현한다"={1, 3, 5}, 사건 ="점수가 3을 넘지 않는다"={1, 2, 3}이면 와 의 사귐은 ={1, 3}이다.

사건 와 가 배반이면 그 사귐사건은 반드시 불가능사건()으로 되며 반대로 이면 사건 와 은 배반사건으로 된다.

사건의 합과 사귐연산은 유한개 또는 셀수 있는 개수의 사건들로 확장할수 있는데 실례로 사건 가 있을 때 를 유한합, 를 셀수 있는 합, 를 유한사귐, 를 셀수 있는 사귐이라고 한다.

3. 사건 의 에 대한 차

로 표기한다. 이것은 사건 의 표본점들중에서 사건 에 포함되지 않는 표본점들로 구성되는 새로운 사건을 의미한다(그림 1.1.7). 확률론적언어로 사건 는 발생하지만 는 발생하지 않는다고 말한다.

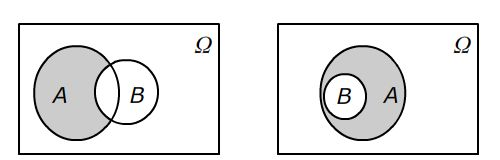


그림 1.1.7

(a) *A*-*B*

(b) *A*-*B* (*A**B*)

한개의 주사위를 던지는 시험에서 사건 ="홀수점수가 출현한다"={1, 3, 5}, 사건 ="점수가 3을 넘지 않는다"={1, 2, 3}이면 의 에 대한 차는 ={5}이다.

가 우연량이라면 다음과 같다.



4. 대립사건

사건 의 대립사건은 로 표시되는데 에는 속하면서 에 속하지 않는 표본점들로 구성된 새로운 사건을 의미한다(그림 1.1.8). 확률론적언어로 "는 발생하지 않는다"고 말할수 있으며 이다. 여기서 의 대립사건은  이고 의 대립사건은  즉 라는것을 알수 있다. 필연사건 와 불가능사건 는 서로 대립사건 즉 이다.

한개의 주사위를 던지는 시험에서 사건 ="홀수점수가 출현한다"={1, 3, 5}의 대립사건은 ={2, 4, 6}, 사건 ="점수가 3을 넘지 않는다"={1, 2, 3}의 대립사건은 ={4, 5, 6}이다.

와 가 서로 대립이기 위한 필요충분조건은 이다.

그림 1.1.8 의 대립사건 







이 성질을 대립사건에 대한 또 다른 정의처럼 리용할수 있다. 즉 사건 와 가 , 를 만족한다면 와 를 서로 대립사건이라고 하고 로 표기한다. 다음과 같은 점에 주의를 돌려야 한다.

① 대립사건은 반드시 배반사건이여야 한다. 즉 . 그러나 배반사건이 반드시 대립사건인것은 아니다.

② 를 로 표시할수 있다.

실례1.1.8 수자 1, 2, …, 9를 반복하여 번()을 뽑는다. 이때 사건 가 "개의 수자를 곱한 적이 10으로 나누인다"라고 하자. 적이 10으로 나누이려면 수자 5와 짝수를 모두 취해야 하기때문에 의 대립사건 는 "뽑혀진 개 수자들가운데 5가 없거나 짝수가 없다"로 된다. 사건 ="뽑혀진 개 수자들가운데 5가 없다", ="뽑혀진 개 수자들가운데 짝수가 없다"라고 하면 이다.

실례1.1.9 를 어떤 우연현상의 세가지 사건이라고 하면

① 사건 "와 는 발생하지만 는 발생하지 않는다"는 로 표시한다.

② 사건 "가운데서 적어도 하나는 발생한다"는 로 표시한다.

③ 사건 "가운데서 적어도 두개의 사건이 발생한다"는 로 표시한다.

④ 사건 "가운데서 두개의 사건이 발생한다"는 로 표시한다.

⑤ 사건 "가 동시에 발생한다"는 로 표시한다.

⑥ 사건 "는 모두 발생하지 않는다"는 로 표시한다.

⑦ 사건 "가 모두 발생하지는 않는다"는 로 표시한다.

5. 사건연산의 성질

- 바꿈법칙

.

- 묶음법칙

.

.

- 분배법칙

.

.

- 쌍대법칙(드모르간공식)

합사건의 대립사건은 대립사건의 사귐과 같다.

.

사귐사건의 대립사건은 대립사건의 합과 같다.

.

사건연산의 쌍대법칙은 매우 쓸모있는 공식이며 이 성질을 증명하기는 어렵지 않다. 여기서 우리는 모임론의 언어를 리용하여 식 (1.1.6)을 증명한다.

식 (1.1.6)의 증명.

이제  즉 라고 하자. 이로부터 가 에도 속하지 않고 에도 속하지 않으며 따라서 와 가 동시에 성립하고 결국 와 가 성립한다. 그러므로 . 이로부터 .

반대로 라고 하자. 그러면 이고 따라서 이며 이것은 가 와 가운데 어느 하나에도 속하지 않는다는것을 의미한다. 즉 이고 이다. 이로부터 .

우의 두가지 결과로부터 이고 식은 증명되였다.

쌍대법칙은 여러 개의 사건과 셀수 있는 사건의 경우에로 일반화될수 있다.

.

.

### 1.1.7 사건령역

여기서 "사건령역"의 개념을 주는 목적은 다음 절에서 사건의 확률을 정의하기 위해서이다.

사건령역이란 직관적으로 표본공간의 일부 부분모임들과 그우에서의 합, 사귐, 차, 대립과 같은 연산의 결과로 이루어지는 모임족이다. 앞으로 사건령역을 로 표기한다. 여기서 "일부 부분모임들"은 전체 부분모임이 될수도 있고 부분모임들의 일부가 될수도 있다. 이것은 표본공간의 성질을 보고 결정해야 한다.

리산표본공간의 경우 모든 부분모임 전체가 필요한 사건령역을 구성할수 있다. 그러나 련속표본공간의 경우 사건령역을 구성하는것은 쉽지 않다. 례를 들어 표본공간이 실수축의 한 구간일 경우 그 길이를 측정할수 없는 부분모임을 인위적으로 구성할수 있는데 이러한 부분모임을 흔히 가측불가능한 모임이라고 부른다. 이 가측불가능한 모임도 사건으로 본다면 그 사건에 대하여 확률을 론할수 없는데 이것은 바라지 않는 현상이다. 이런 현상을 피하려면 련속표본공간의 모든 부분모임을 사건으로 볼 필요가 없이 "측정할수 있는" 부분모임(**가측모임**이라고 부른다)들만 사건으로 보면 된다.

그러면 우리가 어떤 부분모임에 관심을 가져야 하는가, 다시 말하여 에 어떤 원소들이 존재하여야 하는가 하는 문제가 제기된다.

우선 는 와 를 포함해야 하며 또한 사건들에 앞에서 정의된 여러가지 연산들을 적용하여도 사건이 되여야 한다. 특히 셀수 있는 합과 사귐연산에 대해서도 닫겨야 한다. 총체적으로 는 모임의 연산에 대해 모두 닫겨야 한다. 우리는 다음과 같은 사실을 알고있다.

• 사귐연산은 합과 대립연산을 통하여 실현할수 있다(드모르간공식).

• 차연산은 대립과 사귐연산을 통하여 실현할수 있다().

이렇게 합과 대립이 가장 기본적인 연산이므로 사건령역에 대하여 다음과 같이 정의할수 있다.

정의 1.1.1 는 표본공간, 는 의 일부 부분모임들로 구성된 모임족이다. 이제 에 대하여

① ;

② 이면 그의 대립사건에 대하여 ;

③ 이면 그것들의 셀수 있는 합사건에 대하여 

이 만족된다면 를 **사건령역** 또는 **령역**, **대수**라고 한다.

확률론에서는 (, )을 **가측공간**이라고 하며 가측공간에서만 확률을 정의할수 있다. 이때 에는 모두 확률을 론할수 있는 사건들이 있다.

실례1.1.10 대표적인 사건령역들

① 만일 표본공간이 두개의 표본점만 포함()하고 라고 한다면 사건령역은 이다.

② 만일 표본공간에 개의 표본점이 포함()된다면 사건령역 는 빈모임 , 개의 단위원소모임, 개의 2원소모임, 개의 3원소모임, …과 로 구성되는 모임족이다. 이때 에는 모두 개의 사건이 있다.

③ 만일 표본공간에 셀수 있는 개수의 표본점이 포함()된다면 사건령역 는 빈모임 , 셀수 있는 개수의 단위원소모임, 셀수 있는 개수의 2원소모임, …, 셀수 있는 개수의 원소모임과 로 구성되는 모임족이다. 이때 는 셀수 있는 개수의 셀수 있는 원소모임들로 구성된다.

④ 표본공간이 모든 실수를 포함()할 경우 사건령역 의 원소들을 일일이 라열할수는 없고 대신 하나의 기본모임족으로부터 점차 확장하여 구성된다. 구체적인 과정은 다음과 같다.

• 기본모임족 ="모든 반직선들로 구성된 족" 즉 을 취한다.

• 사건령역의 요구를 리용하여 우선 오른쪽열린 닫긴구간 를 확장하여 받아들인다. 여기서 는 임의의 실수이다.

• 다시 닫긴구간, 단일점모임, 왼쪽연린 닫긴구간, 열린구간을 확장하여 받아들인다.



• 마지막으로 유한개 또는 셀수 있는 개수의 합연산과 사귐연산을 리용하여 실수모임의 모든 유한모임, 셀수 있는 모임, 열린모임, 닫긴모임을 확장하여 받아들인다.

우와 같은 확장을 통하여 얻은 모임들이 모두 사건령역의 정의를 만족한다. 이러한 사건령역 를 **보렐(Borel)사건령역**이라고 부르며 령역들의 매 원소(모임)들을 **보렐모임** 또는 **가측모임**이라고 부르는데 이런 가측모임들은 모두 확률을 론의할수 있는 사건들이다.

정의 1.1.2 (표본공간의 분할) 표본공간 에 대하여 개의 사건 에 대하여 들이 서로 배반이고 이라면 을 표본공간 의 **분할**이라고 한다. 또한 셀수 있는 개수의 서로 배반인 사건 들도 표본공간 의 분할로 될수 있다.

분할은 문제를 단순화할수 있기때문에 확률과 통계에 대한 연구에서 자주 리용된다(구체적으로는 1.4.3의 완전확률공식을 참고). 례하면 TV의 천연색농도 는 중요한 품질지표로서 그 리상값은 으로서 천연색농도가 너무 크거나 작은것은 모두 좋지 못하다. 생산과정에 천연색농도를 정확히 으로 조절하는것은 불가능한 일이다. 천연색농도의 가능한 모든 값을 고찰할 필요가 없기때문에 일반적으로 천연색농도를 수요자가 접수할수 있는 상황에 따라 다음과 같은 등급으로 나눈다(는 일정한 상수).

 (1등품),

(2등품),

(3등품),

(불합격품).

이렇게 천연색농도의 표본공간 를 서로 배반인 4개의 사건으로 나누면 하나의 분할 가 얻어진다. 이때 분할 의 모든 가능한 합과 빈모임 로 구성되는 사건령역만 연구하면 되기때문에 이 사건령역을 분할 에 의하여 얻어진 사건령역이라고 하고 라고 표시한다. 이 사건령역은 개의 서로 다른 사건만 포함하므로 문제에 대한 연구가 쉬워진다.

일반적으로 와 같이 개의 사건으로 구성된 분할에 대하여 그것으로부터 얻어지는 사건령역 은 모두 개의 서로 다른 사건을 포함한다. 분할방법은 보통 일부 문제에서 사건령역을 간단히 하기 위하여 연구된다.

련습문제 1.1

1. 아래와 같은 우연시험의 표본공간을 작성하시오.

① 동전 3개 던지기

② 주사위 3개 던지기

③ 앞면이 나올 때까지 련속 동전을 던지기

④ 주머니속에 검은색, 흰색, 붉은색 공이 각각 하나씩 있는데 먼저 임의로 하나를 꺼내고 다시 넣은 다음 하나를 임의로 꺼내기

⑤ 주머니에 검은색, 흰색, 붉은색 공이 각각 하나씩 있는데 먼저 임의로 하나를 꺼내고 다시 넣지 않은 다음에 하나를 임의로 꺼내기

2. 먼저 동전 한개를 던진 다음 앞면이 나타나면(로 표시) 주사위를 던지고 시험을 정지한다. 만일 뒤면이 나타나면(로 표시) 다시 동전을 던지고 시험을 정지한다. 이 시험의 표본공간 는 무엇인가?

3. 를 3개의 사건으로 하고 다음과 같은 사건을 표시하시오.

① 가 모두 발생하거나 모두 발생하지 않는다.

② 가운데서 한개보다 많이는 발생하지 않는다.

③ 가운데서 두개보다 많이는 발생하지 않는다.

④ 가운데서 적어도 두개가 발생한다.

4. 다음과 같은 사건같기식이 성립하는 조건을 지적하시오.

①  ②  ③ 

5. 를 우연량, 그 표본공간을 , 사건라고 할 때 다음과 같은 사건들을 쓰시오.

①  ②  ③  ④ 

6. 3개의 제품을 검사하여 매 제품을 합격품(0으로 표시)과 불합격품(1로 표시)으로 구분한다. 를 3개의 제품가운데서 불합격품의 수로 한다면 다음의 사건들에 포함되는 표본점을 지적하시오. 

7. 다음의 명제가 성립하는가를 따져보시오.

① 

② 이고 이면 

③ 

④ 

8. 사건 이면 반드시 이 성립하는가를 따져보시오.

9. 다음과 같은 사건들의 대립사건을 쓰시오.

① ="동전 두개를 던지면 모두 앞면이 출현한다"

② ="세번 사격하여 모두 목표물을 명중시킨다"

③ ="4개의 부속품을 가공하여 최소 1개가 합격품이다"

10. 다음의 사건의 연산공식을 증명하시오.

①  ② 

11. 가 사건령역이고 일 때 다음을 증명하시오.

① 

② 유한합 

③ 유한사귐 

④ 셀수 있는 사귐 

⑤ 차연산 

## 1.2 확률의 정의와 그 결정방법

이 절에서는 확률론에서 가장 기본적인 문제의 하나인 확률의 정의와 그 결정방법을 소개한다.확률에 대한 간단하고도 직관적인 표현은 바로 우연사건이 발생할수 있는 가능성의 크기이다. 이에 대해 먼저 다음과 같은 경험적사실을 보도록 하자.

① 우연사건의 발생은 우연성을 가지고있다. 그러나 우연사건의 발생가능성은 그 크기가 다르다. 례하면 주머니속에 크기가 같은 공이 10개 있는데 그중 9개는 검은색이고 1개는 붉은색이라고 할 때 주머니에서 1개의 공을 임의로 꺼낼 때 검은색공을 꺼낼 가능성이 붉은색공을 꺼낼 가능성보다 크다는것은 사람들의 공통된 인식이다.

② 우연사건의 발생가능성은 측정할수 있다. 례를 들면 동전을 던졌을 때 앞면이 나타나는 가능성과 뒤면이 나타나는 가능성은 각각 1/2이다. 축구심판은 동전을 던지는 방법으로 량팀의 주장으로 하여금 동등한 기회를 주어 운동장의 자리를 선택하게 한다.

③ 일상생활에서 사람들은 일부 우연사건의 발생가능성에 대해 흔히 백분률 (0부터 1사이의 한개 수)로 측정한다. 례를 들면 복권을 구매한 후 당첨될수도 있고 당첨되지 않을수도 있지만 당첨가능성의 크기를 당첨률로 측정할수 있다. 선택한 제품이 합격품일수도 있고 불합격품일수도 있으나 제품품질의 좋고나쁨은 불합격품률로 측정할수 있다. 또한 갓난아이가 남자일수도 있고 녀자일수도 있으나 남자출생의 가능성은 남자출생률로 측정할수 있다. 이런 당첨률, 불합격품률, 남자출생률은 모두 확률의 원형이다.

확률론의 발전력사에서 확률의 고전적정의, 기하학적정의, 잦음률정의와 주관적정의가 있었다. 이러한 정의들은 일부 우연현상들에만 부합되였다. 그러면 어떻게 모든 우연현상에 부합되는 확률에 대한 가장 일반적인 정의를 내릴것인가? 1900년 수학자 힐베르트 (Hilbert, 1862-1943)는 극히 적은 몇가지 본질적특성으로부터 출발하여 확률의 개념을 설명하는 방법으로 확률에 대한 공리적정의를 확립할것을 제기하였다. 1933년 꼴모고로브(Kolmogorov, 1903-1987)는 처음으로 확률의 공리적정의를 제시하였는데 이 정의는 력사적으로 존재하던 확률에 대한 여러가지 정의들의 공통적인 특성을 포괄할뿐만아니라 그것들의 제한성과 불명확성을 피하였다. 어떤 우연현상이든지 그 정의의 3가지 공리를 만족시켜야만 그것을 확률이라고 할수 있다. 이 공리적정의는 인차 학계의 공인을 받고 확률론발전사에서 하나의 리정표로 되였으며 그후 확률론은 신속히 발전하였다.

### 1.2.1 확률의 공리적정의

정의 1.2.1 를 표본공간, 를 의 어떤 부분모임들로 구성된 사건령역이라고 하자. 어떤 사건 에 대하여 우에 정의된 실값함수 가 다음과 같은것을 만족한다고 하자.

① 비부성공리: 이면 이다.

② 정칙성공리: 

③ 가법성공리(셀수 있는 개수의 가법에 관한 공리): 만일 들이 서로 배반이면



이때 를 사건 의 **확률**이라고 하고 3원소 를 **확률공간**이라고 한다.

확률의 공리적정의는 확률의 본질을 설명한것이다. 확률은 모임(사건)의 함수이다. 만일 사건령역 우에 주어진 함수가 우의 세가지 공리를 만족시키면 확률이라고 한다. 반대로 만일 이 함수가 세가지 공리 중 어느 하나라도 만족시키지 못하면 확률이 아니다.

공리적정의는 어떻게 확률을 결정해야 하는가는 서술하지 않는다. 력사적으로 공리화정의가 나오기 전에 확률의 빈도적정의, 고전적정의, 기하적정의, 주관적정의 등은 모두 일정한 상황에서 각각의 확률을 결정하는 방법을 가지고있었기때문에 확률의 공리적정의가 있은 후에는 그것들을 확률을 결정하는 방법으로 보는것이 타당하다. 아래에 먼저 확률을 결정하는 고전적방법들에서 많이 리용하는 배렬과 조합의 공식을 소개하고 그다음 확률을 결정하는 방법들을 설명한다.

### 1.2.2 배렬과 조합에 대한 공식

순렬과 조합은 모두 개의 원소에서 임의의 개의 원소를 선택하는 총수를 계산하는 공식인데 그 차이는 원소들사이의 선택순서를 따지지 않으면 조합을 리용하고 그렇지 않으면 배렬을 리용한다는것이다. 원소들사이의 순서를 중시하는가 하는 여부는 실제 문제로부터 판별할수 있다. 례를 들면 두 사람이 서로 악수할 때 순서를 따지지 않는다는것이며 줄을 설 때는 순서를 따져야 한다는것이다.

배렬과 조합에 대한 공식들은 모두 다음과 같은 두가지 계산원리에 기초하여 유도한다.

① 곱하기원리

만일 어떤 일을 단계를 거쳐야 완성할수 있고 첫 단계를 하는데 가지 방법, 두번째 단계를 하는데 가지 방법, ……, 단계를 하는데 가지 방법이 있다고 하면 이 일을 완성하는데 모두 가지 방법이 있다.

례를 들면 도시 ㄱ에서 도시 ㄴ까지는 3갈래의 로선이 있고 도시 ㄴ에서 도시 ㄷ까지는 2갈래 로선이 있다면 ㄱ시에서 ㄴ시를 거쳐 ㄷ시까지 이르는데는 총 갈래의 로선이 있다.

② 더하기원리

만일 어떤 일을 가지의 서로 다른 방법으로 완성할수 있고 첫번째 방법에 가지 완성방법, 두번째 방법에 가지 완성방법, ……, 번째 방법에 가지 완성방법이 있다면 이 일을 완성하는데는 모두 가지 방법이 있다.

례를 들면 도시 ㄱ에서 도시 ㄴ로 가는 교통수단에는 뻐스, 기차, 비행기 세가지 류형이 있고 뻐스는 8개 행선, 기차는 5개 행선, 비행기는 3개 행선이므로 ㄱ시에서 ㄴ시까지 모두 8+5+3=16개 행선을 선택할수 있다.

배렬과 조합의 정의, 그 계산공식은 다음과 같다.

① 배렬

개의 서로 다른 원소들가운데서 임의로 ()개 원소를 뽑아서 한줄로 정렬(원소가 나타나는 순서를 고려)한것을 하나의 **배렬**이라고 하며 이런 배렬의 총수를 라고 한다. 곱하기원리에 따라 첫 원소에 대해 개의 선택방법이 있고 두번째 원소에 대해서는 개의 선택방법, ……, 번째 원소에 대해서는 개의 선택방법이 있기때문에 다음과 같다.



이면 **완전배렬**이라고 하고 라고 표시한다. 분명히 이다.

② 중복배렬

개의 서로 다른 원소들가운데서 매번 하나를 임의로 뽑고 그것을 다시 넣은 후 또 하나를 임의로 뽑는데 이렇게 련속 번 뽑아 얻은 배렬을 **중복배렬**이라고 하는데 이런 중복배렬수는 모두 개 있다. 여기서 는 보다 클수 있다.

③ 조합

개의 서로 다른 원소들가운데서 임의로 ()개 원소를 취하여 한개 조로 묶는다(원소들사이의 순서는 고려하지 않는다). 이것을 **조합**이라고 하며 조합의 총수를  또는 로 표시한다. 곱하기원리에 따르면 이러한 조합의 총수는 다음과 같다.



여기에서 , 이라고 약속한다. 조합은 식 을 만족한다.

④ 중복조합

개의 서로 다른 원소들가운데서 매번 하나를 취하고 다시 넣은 후 또 하나를 취하는데 이렇게 련속 번 반복하여 얻은 조합을 **중복조합**이라고 하며 이런 중복조합의 총수는 이다. 여기서 은 보다 클수도 있다.

중복조합의 총수를 얻는것을 다음과 같이 생각할수 있다. 개의 원소를 그림 1.2.1와 같이 개의 상자(개의 성냥가치로 표시)로 그린다. 만일 째 원소가 한번 취해지면 이 상자에 ""로 표식을 한다. 그림 1.2.1에 대한 설명을 한다면 첫번째 원소는 2번 취해졌고 두번째 원소는 0번, 세번째 원소는 1번, ……, 번째 원소는 3번 취해졌다는것이다. 모두 번 취하기때문에 개의 ""가 있으며 개의 ""가 있다. 이렇게 개의 ""와 개의 ""가운데서 두끝에 있는 ""가 움직이지 않고 개의 ""와 ""가 임의로 움직일수 있다. 서로 다른 배치법마다 고유한 선택방식을 나타내므로 중복조합수는 이 개의 위치에서 임의로 개를 선택하여 ""를 놓는것, 혹은 개의 위치에서 임의로 개를 선택하여 ""를 놓는것과 같다.

한편 와 은 서로 같다.

그림 1.2.1 중복조합 설명도

……

우의 언급된 네가지 종류의 배렬, 조합과 그 계산공식은 확률을 결정하는 고전적인 방법들에서 자주 리용된다. 이때 순서와 중복여부를 잘 판단하여 리용해야 한다.

### 1.2.3 확률결정의 잦음률방법

확률을 결정하는 잦음률방법은 대량의 중복실험과정에 잦음률의 안정값으로 확률을 얻는 한가지 방법으로서 그 기본원리는 다음과 같다.

① 고찰되는 사건 와 관계되는 우연현상을 대량으로 반복진행할수 있다.

② 차례의 중복실험에서 를 사건 가 발생한 차수라고 하면 를 사건 의 **빈도수**라고도 부른다. 이때



를 사건 가 발생하는 **잦음률**이라고 한다.

③ 오랜 기간의 실천과정에 사람들은 실험이 반복되는 회수 이 증가함에 따라 잦음률 는 어떤 상수 부근에서 안정된다는것을 알게 되였는데 이 상수를 우리는 **잦음률안정값**이라고 한다. 이 잦음률안정값이 바로 우리가 구하려는 확률이다.

주의할것은 확률을 결정하는 잦음률방법은 합리적이긴 하지만 이 방법의 결점도 매우 뚜렷하다. 현실세계에서는 사람들의 실험을 무한히 반복할수 없고 따라서 잦음률의 안정값을 정확히 얻는것은 불가능하다. 잦음률방법은 확률에 대하여 상상할만한 구체적인 값을 제공하며 실험반복차수 이 비교적 클 때 잦음률을 리용하여 확률에 대한 한가지 근사값을 주게 된다. 통계학에서는 바로 잦음률을 확률의 추정값라고 한다. 례를 들면 축구경기에서 사람들은 벌차기를 성공시킬 가능성에 대해 매우 관심을 가진다. 1930년부터 1988년사이에 세계각지에서 진행된 53,274번의 주요 축구경기에서 15,382개의 벌차기가 있었는데 그중 11,172개의 득점이 이루어졌다. 이로부터 벌차기성공률의 추정값은 11,172/15382=0.726이다.

잦음률법에 의해 결정된 확률이 공리적정의를 만족시칸다는것을 쉽게 검증할수 있다. 비부성과 정칙성은 분명하므로 가법성을 따져보자. 와 가 서로 배반일 때 의 출현수는 의 출현수와 의 출현수를 구하여 더하면 된다. 즉 이고



실례 1.2.1 잦음률의 안정성을 보여주는 몇가지 실례를 설명한다.

① 동전던지기실험([3])

지금까지 많은 사람들이 동전던지기실험을 하였는데 그 결과를 표 1.2.1에 주었다. 표에서 알수 있는바와 같이 앞면이 나타나는 잦음률은 점차 0.5에서 안정된다. 잦음률에 의한 확률결정방법에 의하면 앞면이 나타날 확률은 0.5이다.

표 1.2.1 동전던지기실험의 몇가지 결과

|  |  |  |  |
| --- | --- | --- | --- |
| 실험자 | 동전던진 회수 | 앞면출현 회수 | 잦음률 |
| A | 2048 | 1061 | 0.5181 |
| B | 4040 | 2048 | 0.5069 |
| C | 10000 | 4979 | 0.4979 |
| D | 12000 | 6019 | 0.5016 |
| E | 24000 | 12012 | 0.5005 |

② 영어자모의 잦음률([3])

사람들은 생활실천에서 영어 일부 자모의 출현잦음률이 다른 일부 글자보다 높다는것을 인식하고있다. 그런데 26개의 영어자모가 각각 얼마나 자주 나타날까? 어떤 사람이 여러 대표적인 영문간행물에서 자모의 출현잦음률을 통계해본 결과 매 문자의 리용잦음률이 상당히 안정하다는것을 발견하였다(표 1.2.2). 이 연구는 편리한 곳에 리용잦음률이 가장 높은 자모건을 배렬하는 콤퓨터건반의 설계와 그런 자모들을 짧은 코드로 배렬하는 정보의 부호화 등에서 널리 리용된다.

표 1.2.2 영어자모의 리용잦음률

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 자모 | 리용잦음률 | 자모 | 리용잦음률 | 자모 | 리용잦음률 |
| E | 0.1268 | L | 0.0394 | P | 0.0186 |
| T | 0.0978 | D | 0.0389 | B | 0.0156 |
| A | 0.0788 | U | 0.0280 | V | 0.0102 |
| O | 0.0776 | C | 0.0268 | K | 0.0060 |
| I | 0.0707 | F | 0.0256 | X | 0.0016 |
| N | 0.0706 | M | 0.0244 | J | 0.0010 |
| S | 0.0634 | W | 0.0214 | Q | 0.0009 |
| R | 0.0594 | Y | 0.0202 | Z | 0.0006 |
| H | 0.0573 | G | 0.0187 |  |  |

③ 녀자아이의 출생잦음률([7])

녀자아이의 출생잦음률을 연구하는것은 인구통계에서 매우 중요하다. 이 문제를 가장 먼저 연구한 사람은 라플라스(Laplace,1749-1827)인데 그는 런던, 베를린과 프랑스 전지역의 많은 인구를 조사한 결과 녀자아이의 출생잦음률은 대략 21/43이라는것을 밝혀내였다. 통계학자 크레이머(Cramer, 1893-1985)는 1935년 스웨리예의 공식통계자료(표 1.2.3)를 리용하여 녀자아이의 출생잦음률이 대략 0.482이라는것을 밝혔다.

표 1.2.3 1935년 스웨리예에서 녀자아이가 월별로 태여난 잦음률

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 윌 | 1 | 2 | 3 | 4 | 5 | 6 |  |
| 갓난아이수 | 7280 | 6957 | 7883 | 7884 | 7892 | 7609 |  |
| 녀자아이수 | 3537 | 3407 | 3866 | 3711 | 3775 | 3665 |  |
| 잦음률 | 0.486 | 0.490 | 0.490 | 0.471 | 0.478 | 0.482 |  |
| 윌 | 7 | 8 | 9 | 10 | 11 | 12 | 1년 |
| 갓난아이수 | 7585 | 7393 | 7203 | 6903 | 6552 | 7132 | 88273 |
| 녀자아이수 | 3621 | 3596 | 3491 | 3391 | 3160 | 3371 | 42591 |
| 잦음률 | 0.477 | 0.486 | 0.485 | 0.491 | 0.482 | 0.473 | 0.4825 |

### 1.2.4 확률결정의 고전적방법

확률을 결정하는 고전적방법은 확률론의 발전력사에서 가장 먼저 연구된 방법이다. 그것은 간단하고 직관적이여서 대량의 반복적인 실험을 할 필요가 없으며 경험적인 사실의 기초에서 고찰되는 사건의 가능성에 대해 론리적인 분석을 한 후 그 사건의 확률을 얻어낸다.

고전적방법의 기본원리는 다음과 같다.

① 해당한 우연현상은 유한개의 표본점(례를 들어 개)만을 포함한다.

② 매 표본점의 발생가능성은 같다(등가능성이라고 부른다). 례를 들어 동일한 동전을 던질 때 "앞면이 출현한다"와 "뒤면이 출현한다"의 가능성, 동일한 주사위를 한개 던질 때 매 점수(1~6)가 출현할 가능성, 한조의 주패에서 임의로 한장을 뽑을 때 매 카드가 뽑히울 가능성들은 서로 같다.

③ 만일 사건 가 개의 표본점을 포함하고 있다면 사건 의 확률은



우의 공식에 의해 결정된 확률은 공리적정의를 만족시킨다. 그중에서 비부성과 정규성은 분명하다. 가법성을 만족하는가를 잦음률방법에서와 류사하게 검증한다. 와 가 서로 배반일 때 의 표본점수는 의 표본점수와 의 표본점수를 구하여 더하면 된다. 즉 가법성 을 만족한다.

고전적방법은 확률론의 발전초기에 확률을 결정하는데 흔히 리용하던 방법이므로 얻은 확률을 고전적확률이라고도 한다. 고전적인 방법으로 사건 의 확률을 구하는것은 에 포함된 표본점의 개수와 에 포함된 표본점의 총수를 계산하는것으로 귀착된다. 때문에 계산에서는 흔히 배렬, 조합을 리용한다.

실례 1.2.2 두개의 동전을 던져서 앞면과 뒤면이 나올 확률을 구하시오.

풀기. 이 문제의 표본공간은 ={(앞면, 앞면), (앞면, 뒤면), (뒤면, 앞면), (뒤면, 뒤면)}이다. 따라서 에 포함된 표본점의 개수는 4 이고 사건 "앞면 하나와 뒤면 하나이 출현한다"에 포함된 표본점의 개수는 2이므로 구하려는 확률은 1/2이다.

만일 표본공간을 ={(두번의 앞면), (두번의 뒤면), (한번의 앞면, 한번의 뒤면)}으로 표시한다면 이 3개의 표본점은 등가능하지 않다.

고전적확률을 계산할 때 일반적으로 표본공간을 상세하게 쓸 필요는 없지만 반드시 표본점들이 등가능해야 한다. 아래에 비교적 쓸모있는 모형을 준다.

실례 1.2.3(표본추출모형) 모두 개의 제품이 있는데 그중 개는 불합격품이고 개는 합격품이다. 개의 제품()을 우연선택하고 사건 ="선택된 개의 제품가운데서 개가 불합격품이다"의 확률을 구하시오. 여기서 이다.

풀기. 먼저 표본공간 의 표본점 총수를 계산한다. 개의 제품가운데서 임의로 개를 취하고 순서를 고려하지 않기때문에 표본점의 총수는 이다. 또한 우연선택하기때문에 이 표본점들은 모두 등가능하다.

먼저 사건 의 확률을 계산한 다음 의 확률을 계산한다.

사건 ="꺼낸 개 제품가운데서 불합격품이 0개이다"="꺼낸 개 제품은 모두 합격품이다"이므로 꺼낸 개 제품은 모두 개의 합격품가운데서 뽑았다는것을 의미한다. 따라서 가지 선택방법이 있으며 따라서 의 확률은 .

사건 ="꺼낸 개 제품가운데서 불합격품이 1개이다"에서 꺼낸 개 제품가운데서 오직 1개만 불합격품이고 나머지 개는 합격품이므로 두 단계를 거쳐야 한다.

1단계: 개의 불합격품중에서 우연적으로 1개를 선택한다. 모두 가지 선택방법이 있다.

2단계: 개의 합격품중에서 우연적으로 개를 선택한다. 모두 가지 선택방법이 있다.

따라서 곱하기원리에 의하여 에는 개의 표본점이 있기때문에 그 확률은

.

에 대한 분석을 통하여 일반적인 사건 에 포함되는 표본점의 개수를 쉽게 계산할수 있다.이 발생하자면 개의 불합격품가운데서 개를 선택하고 다시 개의 합격품가운데서 개를 선택하여야 한다. 곱하기원리에 따라 에는 개의 표본점이 있다. 따라서 의 확률은



여기서 이므로 이여야 하며 그렇지 않을 경우 확률은 0이다.

를 취하면 다음과 같다.



우의 계산결과를 표 1.2.4에 제시하였다.

표 1.2.4 사건 의 확률

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 0 | 1 | 2 | 3 |
|  |  |  |  |  |

표에서 확률들의 합이 1이기때문에 이 0, 1, 2, 3을 취하는 네가지 경우가운데서 반드시 한가지는 발생한다. 그러므로 **확률분포**라고 할수 있다. 만일 을 우연량으로 보면 이 분포는 의 분포이다.

겉으로 보기에 확률분포는 서로 배반인 사건들과 그 확률을 하나의 우연량으로 관련시켜 구성한 표같지만 사실 확률분포는 우연량이 취하는 값의 확률법칙을 전면적으로, 동적으로 묘사하며 그것으로부터 더 많은 정보를 얻을수 있고 우연현상을 더 깊이 연구할수 있다. 제2장부터는 주로 우연량과 그 확률분포를 론의한다.

실례 1.2.4(반환표본화) 표본화에는 비반환표본화와 반환표본화의 두가지가 있다. 우의 례에서는 비반환표본화에 대하여 론의하였다. 반환표본화는 한개을 선택한 다음 다시 반환하고 또 다시 한건을 선택하여 개를 선택할 때까지 반복한다. 이 실례에서는 반환표본화의 경우에 사건 ="꺼내여진 개의 제품가운데 개의 불합격품이 있다"의 확률을 구한다.

풀기. 먼저 표본공간 의 표본점 총수를 계산한다. 처음 뽑을 때에는 개가운데서 아무거나 하나를 뽑을수 있으며 개의 선택방법이 있다. 다시 반환하고 뽑기때문에 두번째 선택할 때에도 역시 가지 선택방법이 있다. 이렇게 하면 매번 가지 방법으로 뽑고 모두 번 뽑기때문에 총 가지 가능한 표본점이 있다.

사건 ="꺼낸 개의 제품은 모두 합격품이다"가 발생하려면 개의 합격품가운데서 반환표본화를 번 해야 하므로 에는 개의 표본점이 있고 따라서 의 확률은

.

사건 ="꺼낸 개 제품가운데 불합격품이 한개 있다"가 발생하려면 반드시 개의 합격품가운데서 반환표본화를 번, 개의 불합격품가운데서 한번 추출하여야 한다. 따라서 가지 선택법이 있다. 이 불합격품이 처음으로 뽑히울수도 있고 두번째로 뽑히울수도 있으며 번째로 뽑히울수도 있다. 이렇게 가지 가능성이 있으므로 은 개의 표본점이 있다. 따라서 의 확률은 .

사건 ="뽑힌 개의 제품가운데 개의 불합격품이 있다"이 발생하려면 개의 합격품가운데서 반환표본화를 번 하고 개의 불합격품가운데서 반환표본화를 번 해야 하므로 가지 선택법이 있다. 이 개의 불합격품이 번중의 선택중 임의로 번만 선택되면 되므로 가지 가능성이 있다.

그러므로 사건 은 개의 표본점을 가지며 그 확률은



반환표본화를 진행하므로 불합격품이 전체 제품에서 차지하는 비률 은 변하지 않는다. 이 비률을 로 표시하면 식 (1.2.7)을 다음과 같이 쓸수 있다.



마찬가지로 로 취하면



우의 계산결과를 표 1.2.5에 제시하였다.

표 1.2.5 사건 의 확률

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 0 | 1 | 2 | 3 | 4 |
|  |  |  |  |  |  |

표에서 확률들의 합이 1이므로 이것도 역시 확률분포이다. 또한

.

실례 1.2.5(복권문제) 일종의 복리복권을 "35-7선택"이라고 하는데 구매할 때 01, 02, …, 35가운데서 임의로 7개의 번호를 선택하며 추첨할 때 01, 02, …, 35가운데서 중복되지 않게 기본번호 7개와 특수번호 하나를 선택한다. 당첨규칙은 다음과 같다.

표 1.2.6 "35-7선택"의 당첨규칙

|  |  |
| --- | --- |
| 당첨등급 | 당첨규칙 |
| 1등 | 7개의 기본번호가 다 맞는다. |
| 2등 | 6개의 기본번호와 특수번호가 맞는다. |
| 3등 | 6개의 기본번호가 맞는다. |
| 4등 | 5개의 기본번호와 특수번호가 맞는다. |
| 5등 | 5개의 기본번호가 맞는다. |
| 6등 | 4개의 기본번호와 특수번호가 맞는다. |
| 7등 | 4개의 기본번호 또는 3개의 기본번호와 특수번호가 맞는다. |

매 당첨등수의 당첨확률을 구하시오.

풀기. 번호를 중복되지 않게 선택하는것은 비반환표본화이므로 표본공간 는 개의 표본점을 포함한다.

당첨되려면 세가지 류형의 번호를 뽑아야 한다. 1류형은 7개의 기본번호, 2류형은 1개의 특수번호, 3류형은 27개의 무효번호이다. 실례 1.2.3에서는 두 종류의 원소(합격품과 불합격품)들가운데서 뽑았는데 이 실례에서는 세 류형의 번호에서 뽑는다. 를 등에 당첨될 확률(=1, 2, …, 7)이라고 하면 실례 1.2.3에서와 마찬가지 방법으로 매 등수에 당첨될 확률을 구할수 있다.







사건 "당첨된다"를 로 표시하면 는 사건 "당첨되지 않는다"이고 또한 이므로



이것은 100명중 약 3명이 당첨되며 1등에 당첨될 확률은 으로서 2천만명중 3명정도밖에 1등에 당첨되자 않는다는것을 말해준다.

실례1.2.6(상자모형) 개의 공으로 있는데 매 공은 개의 다른 상자중 어느 하나에 들어갈수 있으며 매 상자에 넣어지는 공의 개수는 제한되지 않는다.

① 지정한 개의 상자에 각각 구가 한개씩 있을 확률 ;

② 개의 상자에 각각 구가 한개씩 있을 확률

풀기. 매 구를 개의 상자중 임의의 하나에 넣을수 있기때문에 개의 구를 넣는 방법은 모두 가지가 있으며 모두 등가능하다.

① 각각 하나의 공이 들어있는 개의 상자가 지정되였기때문에 나머지 공이 없는 개의 상자도 동시에 지정된다. 따라서 이 지정된 개의 상자에 개의 공이 각각 하나씩 들어가는 경우의 수를 고려하기만 하면 된다. 첫번째 공은 가지 방법이 있고 두번째 공은 가지방법, ……, 번째 공은 한가지 방법밖에 없다. 그러므로 곱하기원리에 의하여 가능한 총수는 이며 그 확률은



② ①와의 차이는 이 개의 상자를 개의 상자가운데서 임의로 선택할수 있다는것이다. 이때 두 단계로 나누어 진행할수 있다.

첫번째 단계는 개의 상자중에서 공을 넣을 개의 상자를 임의로 선택하는것이다. 이때 가지 방법이 있다. 두번째 단계로 개의 공을 선택된 개의 상자에 하나씩 넣는다. 이것은 가지 방법이 있다. 따라서 곱하기원리에 의하여 가능한 총수는 가지이다.

사실 이것을 더 직접적으로 생각할수 있다. 첫번째 공은 개의 상자중 임의의 하나에 넣을수 있고 두번째 공은 나머지 개의 상자중의 임의의 하나에, ……, 번째 공은 나머지 개의 상자중 임의의 하나에 넣을수 있는데 곱하기원리에 따라 같은 수를 얻을수 있다. 따라서 그 확률은



상자모형은 공과 상자 문제를 론하는것처럼 보이지만 실제로 이 모형을 많은 실천문제들에 적용할수 있다.

확률론에서 유명한 "생일문제"를 상자모형으로 론의해보자.

실례 1.2.7(생일문제) 명의 생일이 모두 다를 확률 은 얼마인가?

풀기. 명의 사람을 개의 공처럼 보고 1년 365일을 개의 상자처럼 본다면 "명의 생일이 모두 다르다"는것은 "개의 모든 상자에 각각 한개의 공이 있다"와 같으므로 사람 명의 생일이 모두 다를 확률은 다음과 같다.



웃식은 간단해보이지만 구체적인 계산은 시끄럽기때문에 다음과 같은 방법으로 근사계산을 할수 있다.

① 이 작을 때 식 (1.2.10)의 오른변에서 매 인수의 두번째 마디사이의 적 는 모두 무시할수 있으므로 다음과 같은 근사공식을 얻는다.



② 가 비교적 큰 때 작은 정수 에 대하여 이므로 식 (1.2.10)으로부터



례컨대 일 때 식 (1.2.12)에 의한 근사값은 0.8840 이고 정확한 값은 이다. 일 때 근사값은 0.3073이고 정확한 값은 이다.

이 결과는 보면 사람들의 일반적리해와 차이난다. 왜냐하면 많은 사람들은 1년이 365일이므로 30명의 생일이 모두 다를 가능성이 비교적 크며 적어도 1/2보다는 클것이라고 생각하기때문이다. 심지어 100명의 생일이 모두 다를 가능성도 크다고 보는 사람도 있다. 다른 값에 대하여 근사공식 (1.2.12)으로 계산한 값을 표 1.2.7에 보여준다.

표 1.2.7 의 근사값

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | 10 | 20 | 30 | 40 | 50 | 60 |
|  | 0.8840 | 0.5942 | 0.3037 | 0.1180 | 0.0349 | 0.0078 |
|  | 0.1160 | 0.4058 | 0.6963 | 0.8820 | 0.9651 | 0.9922 |

표의 마지막 행은 대립사건 "명가운데 적어도 두 사람의 생일이 같다"의 확률 이다. 일 때  즉 60명의 사람들가운데서 적어도 두 사람의 생일이 같을 확률이 99%를 초과한다는것을 말한다. 또한 일 때 라는 결론을 얻을수 있다.

### 1.2.5 확률결정의 기하방법

확률을 결정하는 기하방법의 기본원리는 다음과 같다.

① 만일 우연현상의 표본공간 가 어떤 구역을 꽉 채우면 그 척도(길이, 면적, 체적 등)의 크기는 로 표시한다.

② 임의의 점이 척도가 같은 부분구역에 놓인다고 할 때 그 위치는 다를수 있지만 가능성은 같다. 례를 들어 표본공간 에서 단위 정사각형 와 직각변이 1, 2인 직3각형 가 있을 때 두 구역의 면적이 같으므로 구역과 구역에 놓일 확률은 서로 같다(그림 1.2.1).

③ 만일 사건 가 의 한 부분구역(그림 1.2.2)이고 척도를 로 표시하면 사건 의 확률은



이 확률을 **기하확률**이라고 하는데 확률의 공리적정의를 만족시킨다.

기하확률을 구하는데서 중요한것은 표본공간 와 구하는 사건 를 도형(일반적으로 평면이나 공간도형을 리용한다)으로 잘 묘사하는것이다. 그리고 관련도형의 척도(일반적으로 면적, 체적)를 계산한다.



*A*

1

1

1

2

*B*

*A*



그림 1.2.1 척도가 같은 부분구역에 놓이는것들의 등가능성

그림 1.2.2 기하확률

실례1.2.8(면회문제) 두 사람 a, b가 오후 6시부터 7시사이에 어떤 장소에서 만나는데 먼저 온 사람은 다른 한 사람을 20분 기다리다가 시간이 지나면 떠나기로 약속하였다. 두 사람이 만날 확률을 구하시오.

풀기. 와 로 각각 두 사람 a, b가 약속한 장소에 도착한 시간(분단위)을 표시하고 평면우에 직각자리표계를 구성한다(그림 1.2.3). 와 는 모두 0~60분 이내에 도착할수 있기때문에 등가능성으로부터 하나의 기하확률문제라는것을 알수 있다. 의 모든 가능한 값은 한변의 길이가 60인 바른4각형구역내에 존재하며 그 면적은 이다.

그리고 사건 ="두 사람이 만난다"는 에 해당하므로 그림의 어두운 부분이고 그 면적은 , 식 (1.2.13)으로부터 .

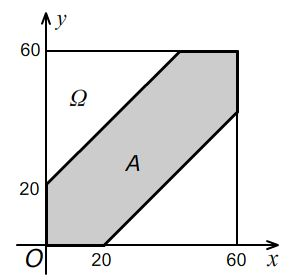


그림 1.2.3 면회문제에서의 와 

결과 이 규칙대로 할 때 두 사람이 만날 확률은 0.6을 넘지 않는다. 만나는 시간을 오후 6시부터 6시 30분까지로 바꾸고 나머지를 그대로 할 경우 두 사람이 만날 확률은 0.8889로 높아진다.

실례 1.2.9(Buffon의 바늘문제([1] 참고)) 평면에 간격이 인 평행선을 그려놓고 길이가 인 바늘을 평면에 던져 바늘이 임의의 평행선과 사귈 확률을 구하시오.

풀기. 그림 1.2.4와 같이 로 바늘의 가운데점과 가장 가까운 평행선사이의 거리를 표시하고 로 바늘과 이 직선사이의 끼움각을 표시한다. 표본공간 는 를 만족시킨다.

이 두 식으로부터 평면우의 직사각형 를 확정할수 있는데 이것이 바로 표본공간이며 그 면적은 이다. 이때 바늘과 평행선과 사귀는(사건 로 표시) 필요충분조건은 이다. 이 부등식으로 표시되는 구역은 그림 1.2.5의 어두운 부분이다.

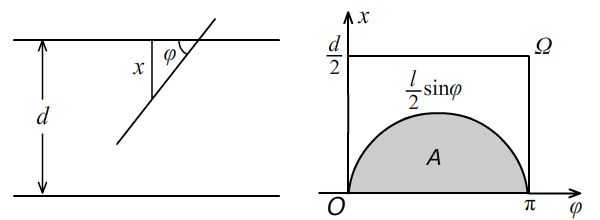


그림 1.2.4 Buffon의 바늘문제

그림 1.2.5 Buffon의 바늘문제에서 와 

바늘을 평면에 임의로 던졌기때문에 등가능성으로부터 이것은 하나의 기하확률문제이라는것을 알수 있다. 따라서 .

만일 를 안다면 웃식으로부터 의 값을 계산할수 있다. 반대로 의 값을 안다면 웃식을 리용하여 를 구할수 있다. 이때 의 값은 실험에서 얻은 잦음률로 계산할수 있다. 즉 바늘을 번 던지고 그가운데서 바늘과 평행선이 번 사귀면 잦음률 을 의 추정값으로 할수 있다. 그리하여 로부터 을 얻을수 있다. 오래전에 일부 학자들이 직접 이 실험을 하였는데 표 1.2.8에 그들의 실험결과가 기록되여있다.

표 1.2.8 Buffon의 바늘실험결과

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 실험자 | 년도 |  | 던지기차수 | 사귄차수 | 의 근사값 |
| A | 1850 | 0.8 | 5000 | 2532 | 3.1596 |
| B | 1884 | 0.75 | 1030 | 489 | 3.1595 |
| C | 1901 | 0.83 | 3408 | 1808 | 3.1415 |
| D | 1925 | 0.54 | 2520 | 859 | 3.1795 |

이것은 기묘한 방법으로서 하나의 우연실험을 설계하여 한 사건의 확률과 어떤 미지수를 련관시키고 반복실험을 통하여 잦음률로 확률을 추정하면 미지수의 근사풀이를 얻을수 있다. 일반적으로 실험차수가 많을수록 구하는 근사풀이가 더 정확하다. 콤퓨터의 출현과 함께 사람들은 그것을 리용하여 설계한 우연실험을 대량으로 반복모의할수 있게 되였다. 이런 방법이 빨리 발전하고 날리 리용되였는데 이것을 **우연모의법** 또는 **몬테카를로법**(Monte Carlo method)이라고 한다.

실례1.2.10 길이가 인 선분에서 두 점을 임의로 취하여 세 부분으로 나누고 그것들이 3각형을 이룰 확률을 구하시오.

풀기. 선분을 임의의 세 부분으로 나누기때문에 등가능성으로부터 이것은 기하확률문제임을 알수 있다. 분할된 세 부분의 길이를 각각 로 표시(그림 1.2.6)하는데 그것들사이에 다음과 같은 관계가 성립한다. .

세번째 식은 과 동일하다. 따라서 표본공간은 다음과 같다(그림 1.2.7 참고).

. 의 면적은 과 같다.

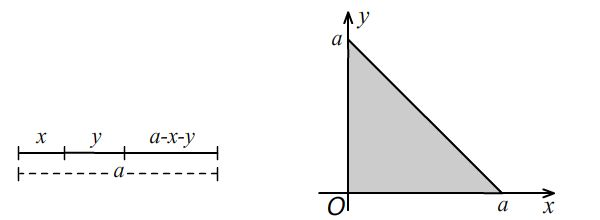


그림1.2.6 길이가 인 선분의 분할

그림1.2.7 선분을 세분할하는 표본공간 

그리고 3각형의 구성조건에 따라 3각형의 임의의 두 변의 합이 세번째 변보다 크므로 사건 에 포함되는 표본점은 다음 식을 만족하여야 한다.



정리하면 . 따라서 사건 는 그림 1.2.8의 어두운 부분으로 표시된다. 사건 의 면적은 이고 결국 .

### 1.2.6 확률을 결정하는 주관적방법

현실세계에서 일부 우연적인 현상은 중복될수 없거나 대량으로 중복될수 없는데 이때 사건의 확률을 어떻게 결정하는가?

통계학계의 베이스학파는 사건의 확률은 사람들이 경험에 근거하여 그 사건의 발생가능성에 대하여 주는 개인적인 관념이라고 여긴다. 이렇게 주어지는 확률을 **주관적확률**이라고 한다.

이처럼 경험을 리용하여 우연사건의 발생가능성 크기를 결정하는 실례는 매우 많으며 사람들은 흔히 일부 주관적확률에 따라 일을 처리한다.

실례 1.2.12 주관적인 방법으로 확률을 결정하는 실례.

① 기상예보에서는 흔히 "래일 비가 올 확률이 90%이다"고 말한다. 이것은 기상전문가들이 기상관련전문지식과 최근의 기상상황에 기초하여 제시하는 주관적인 확률이다. 이 소식을 들은 사람들은 밖에 나갈 때 대부분 우산을 가지고 나간다.

② 어떤 기업가는 다년간의 경험과 당시의 시장정보에 근거하여 "어떤 새제품이 앞으로 시장에서 잘 팔린다"의 가능성은 80%라고 생각한다.

③ 한 외과의사는 자기의 다년간의 림상경험과 환자의 병세에 근거하여 "이 수술이 성공한다"의가능성이 90%라고 인정한다.

④ 한 교원은 자기의 다년간의 교수경험과 두 학생 a, b의 학습정형에 근거하여 "학생 a가 대학에 입학한다"의 가능성은 95%, "학생 b가 대학에 입학한다"의 가능성은 40%라고 생각한다.

우의 실례로부터 다음과 같은것을 알수 있다.

① 주관적확률과 주관적억측은 본질적으로 다르다. 주관적확률은 당사자가 고찰되는 사건에 대한리해가 깊고 풍부한 경험이 있어야 하며 심지어 이 분야의 전문가여야 하며 과거의 정보와 당시의 정보에 대해 자세하게 분석할수 있어야 한다. 이렇게 결정한 주관적확률은 믿음성이 있다. 어떤 의미에서 말하면 이런 풍부한 경험을 리용하지 않는것도 일종의 랑비라고 할수 있다.

② 주관적인 방법으로 얻어낸 우연사건의 발생가능성크기는 본질적으로 우연사건의 확률에 대한 추측과 예측이다. 결론의 정확성은 실천에서 검증하고 수정해야 하지만 결론의 신뢰성은 통계적의미에서 그 가치가 있다.

③ 우연적인 현상이 많이 반복될수 없을 때에는 주관적인 방법으로 결심과 판단을 내리는것이 적합하다. 이 점에서 볼 때 주관적방법은 잦음률방법의 한가지 보충이다.

이밖에 설명해야 할것은 주관적확률의 결정에 있어서 자기의 경험에 근거할수도 있고 다른 사람의 경험도 리용할수 있다는것이다. 주관적확률은 공리적정의에 부합되여야 한다.

련습문제 1.2

1. 조합수 에 대하여 다음의 식들을 증명하시오.

① 

② 

③ 

④ 

⑤ 

⑥ .

2. 동전 3개를 던져 앞면이 적어도 하나 나올 확률을 구하시오.

3. 정의 옹근수 2개를 임의로 취할 경우 그 합이 짝수가 될 확률을 구하시오.

4. 주사위를 2개 던져 다음과 같은 사건의 확률을 구하시오.

① 점수의 합이 6이다.

② 점수의 합이 6을 초과하지 않는다.

③ 6점이 적어도 한개 있다.

5. 1원 2차방정식 에서 는 주사위를 련이어 두번 던져 나타나는 점수이다. 이 방정식이 실수뿌리를 가질 확률 와 중복뿌리를 가질 확률 를 구하시오.

6. 52장의 주패장들가운데서 임의로 4장을 뽑을 때 아래 사건들의 확률을 구하시오.

① 모두 스페이드이다.

② 무늬가 같다.

③ 무늬와 색갈이 같은 두장이 없다.

④ 색갈이 같다.

7. 9개의 제품가운데서 2개가 불합격품이라고 하자. 반환하지 않고 임의로 2개를 선택할 때 그것들이 모두 합격품일 확률, 2개 중 1개만 합격품일 확률, 합격품이 없을 확률은 각각 얼마인가.

8. 주머니에 흰 공 7개, 검은 공 3개가 있을 때 임의로 선택한 두개 공을 선택하였다.

① 두개 공의 색갈이 같을 확률을 구하시오.

② 두개 공의 색갈이 서로 다를 확률을 구하시오.

9. 주머니 a에는 흰 공 5개, 검은 공 3개가 있고 주머니 b에는 흰 공 4개, 검은 공 6개가 있다. 두개의 주머니에서 각각 한개의 공을 취할 때 두개 공의 색갈이 같을 확률을 구하시오.

10. 개의 수 1, 2, …, 중에서 임의의 2개를 골라 그가운데서 하나가 보다 작고 다른 하나가 보다 클 확률을 구하시오.

11. 주머니속에 10개의 공이 있는데 1부터 10까지 번호가 매겨져있다. 그중 4개를 반환하지 않고 선택하여 그 공들의 번호를 적는다. 이때

① 최소번호가 5일 확률을 구하시오.

② 최대번호가 5일 확률을 구하시오.

12. 주사위 3개를 던질 때 다음과 같은 사건의 확률을 구하시오.

① 얻은 최대점수가 5보다 같거나 작다.

② 얻은 최대점수가 5이다.

13. 책 10권을 임의의 책꽂이에 꽂을 때 그중 지정한 4권이 한곳에 꽂일 확률을 구하시오.

14. 명의 사람이 우연적으로 원탁에 둘러앉을 때 두 사람 a, b가 서로 나란히 앉을 확률을 구하시오.

15. 동시에 5개의 주사위를 던져 점수를 관측한다. 아래의 결과를 증명하시오.

① (매 점수가 모두 다르다)=0.0926

② (한쌍의 점수가 같다)=0.4630

③ (두쌍의 점수가 같다)=0.2315

④ (3개 점수가 같다)=0.15154

⑤ (4개 점수가 같다)=0.019

⑥ (5개 점수가 같다)=0.0008.

16. 한 사람이 풀 6대를 손에 쥐고 머리부와 꼬리부만 드러낸 다음 우연적으로 머리부 6개를 2개씩 이어놓고 6개의 꼬리부도 2개씩 이어놓는다. 6개의 풀대들이 하나의 고리로 련결될 확률을 구하시오.

17. 개의 "0"과 개의 "1"을 임의로 배렬할 때 두개의 "1"이 나란히 있지 않을 확률을 구하시오.

18. 10개의 제품중 2개가 불합격품이다. 그중 4개를 임의로 취할 때 불합격품수 의 확률분포를 구하시오.

19. 명의 남자아이, 명의 녀자아이()를 임의로 한 줄로 세울 때 두 녀자아이가 서로 나란히 서지 않을 확률을 구하시오.

20. 4개의 주머니에 공 3개를 임의로 넣을 때 주머니에 있는 공의 최대개수 의 확률분포를 구하시오.

21. 12개의 공을 3개의 상자에 임의로 넣고 첫 상자에 공이 3개 있을 확률을 구하시오.

22. 완전히 같은 개의 구를 개의 상자에 임의로 넣을 때 다음과 같은 확률을 구하시오.

① 어떤 지정된 상자에 개의 구가 있을 확률.

② 개의 빈 상자가 있을 확률.

③ 어떤 지정된 개의 상자에 개의 공이 들어있을 확률.

23. 구간 (0,1)에서 임의의 두 수를 취할 때 두 수의 합이 7/5보다 작을 확률을 구하시오.

24. 두척의 기선 a, b 가 동시에 두척을 정박시킬수 없는 부두를 향하여 항행하고 그것들이 하루동안에 도착하는 시간은 등가능하다. 만일 a의 정박시간이 1시간이고 b의 정박시간이 2시간이라면 그것들중 어느 한척도 부두가 비기 기다리지 않을 확률을 구하시오.

25. 평면에 간격이 인 평행선을 그리고 변의 길이가 (모두 보다 작다)인 3각형을 그 평면에 임의의 방향으로 던질 때 3각형과 평행선이 사귈 확률을 구하시오.

26. 반경이 인 원안에 활줄을 그린다. 이것과 그에 수직인 직경이 사귀는 점의 위치는 그 직경에서 모두 등가능하다. 즉 사귐점이 직경의 어떤 구간안에 놓일 가능성과 그 구간의 길이는 비례한다. 이때 활줄의 길이가 보다 클 확률을 구하시오.

27. 한 질점이 평면에서 축, 축과 직선 로 둘러싸인 3각형안에 놓인다고 가정하고 이 3각형안의 매 점에 떨어질 확률은 같다고 하자. 즉 이 3각형안의 임의의 구역에 떨어질 확률은 이 구역의 면적에 비례한다. 질점의 위치가 를 만족할 확률은 얼마인가?

28. 이라고 하고 임의의 두개 수 에 대하여 라고 하면 일 확률을 구하시오.

29. 주관적인 방법으로 대학생들가운데서 안경을 낄 확률을 결정하시오.

30. 주관적인 방법으로 학생들가운데서 시험에서 부정행위를 할 확률을 결정하시오.

## 1.3 확률의 성질

아래에서 확률의 공리적정의(비부성, 정칙성, 가법성)를 리용하여 확률의 일부 성질들을 유도한다.

우선 확률의 정칙성에서 필연사건 의 확률이 1이라는것을 설명하였다. 따라서 불가능사건 의 확률은 0이여야 하며 아래에서 이 성질을 증명한다.

성질 1.3.1 이다.

증명. 임의의 사건과 불가능사건의 합은 여전히 그 사건 자체이므로 .

불가능사건과 임의의 사건은 서로 배반이므로 가법성공리에 의하여 . 또한 이므로 . 따라서 비부성공리에 의하여 .

### 1.3.1 확률의 가법성

확률의 가법성은 서로 배반인 사건 들에 대하여 그의 셀수 있는 합의 확률은 개별적으로 확률을 구하여 합하는 방법으로 얻을수 있다는것을 밝혔다. 그러면 유한개의 서로 배반인 사건 에 대하여 그의 유한합의 확률도 개별적으로 확률을 각각 구한후 더할수 있는가? 아래에 이 문제에 대한 대답을 준다.

성질 1.3.2(유한가법성) 유한개의 사건 이 서로 배반이면



증명. 이제 셀수 있는 개수의 에 대하여 셀수 있는 개수의 가법성을 적용하면



따라서 결론이 증명된다.

유한가합성을 리용하여 대립사건의 확률을 구하는 공식을 얻을수 있다.

성질 1.3.3 임의의 사건 에 대하여

.

증명. 와 는 서로 배반이고 이므로 확률의 정규성과 유한가법성으로부터 . 따라서 을 얻는다.

어떤 사건은 직접 고려하는것은 비교적 복잡하지만 그 대립사건을 고려하는것은 비교적 간단하다. 이러한 종류의 문제는 아래에 제시한 실례와 같이 성질 1.3.3을 리용할수 있다.

실례 1.3.1 크기와 모양이 같은 36개의 전구가운데서 4개는 6W짜리이고 나머지는 모두 4W짜리이다. 여기서 3개를 임의로 취할 때 적어도 하나가 6W짜리일 확률을 구하시오.

풀기. 사건 "꺼낸 세개가운데서 적어도 한개가 6W이다"를 로 표시하면 는 세가지 경우를 포함한다. 즉 하나는 6W, 두개는 4W이거나 두개가 6W, 한개가 4W, 또는 세개가 모두 6W인 경우이다. 대신 의 대립사건 는 하나의 경우 즉 "선택된 3개는 모두 4W이다"만 포함하는데 실례1.2.3의 표본화모형으로부터 . 그러므로 

실례 1.3.2 같은 동전을 5번 던져 앞면과 뒤면이 다 나올 확률을 구하시오.

풀기. 사건 가 "5번 동전을 던져 앞면과 뒤면이 다 나온다"라고 하면 앞면이 출현하는 회수가 1~4번이 될수 있기때문에 의 경우가 비교적 복잡하다. 그러나 의 대립사건 은 "5번 모두가 앞면이다"(로 표시한다)이거나 "5번 모두가 뒤면이다"(로 표시한다)이므로 상대적으로 간단하다. 즉 이고 와 는 서로 배반이다. 그러므로 대립사건공식과 확률의 유한가법성으로부터



### 1.3.2 확률의 단조성

사건 가 에 포함될 때(즉 의 발생이 의 발생을 초래한다) 사건 가 보다 쉽게 발생한다는것을 알수 있다. 그러면 의 확률이 의 확률보다 커야 한다. 이것은 다음과 같은 성질 1.3.4로 설명할수 있다.

성질 1.3.4는 포함관계에 있는 두 사건의 차사건에 대한 확률공식을 주며 성질 1.3.5는 임의의 두 사건의 차사건의 확률공식을 준다.

성질 1.3.4 만일 라면

.

증명. 이므로 이고 와 는 서로 배반이다. 유한가법성으로부터 이므로 을 얻게 된다. 이렇게 결론이 증명된다.

따름(단조성) 이면 .

우의 따름의 거꿀명제 즉 이면 가 성립하지 않는다는것을 반례를 들어 쉽게 알수 있다.

성질 1.3.5 임의의 두 사건 , 에 대하여

.

증명. 이고 이므로 성질 1.3.4로부터 .

성질 1.3.4을 리용하여 비교적 복잡한 사건의 확률을 구할수 있다.

실례 1.3.3 주머니에 번호가 인 개의 구가 있는데 그중에서 반환을 하면서 임의로 번 꺼낼 때 개 공의 최대번호가 일 확률을 구하시오.

풀기. 사건 가 "꺼낸 개의 공의 최대번호는 이다"라고 하자. 그런데 "최대번호가 "인것은를 1번 꺼낼수도 있고 두번 꺼낼수도 있으며 번 꺼낼수도 있기때문에 를 직접 고려하는것은 비교적 복잡하다. 그래서 사건 를 "꺼낸 개 공의 최대번호는 ()보다 같거나 작다"라고 하면 의 발생은 번호가 인 공들가운데서 꺼내기만 하면 된다. 결국 고전적확률로부터 임을 알수 있다. 또한 이므로 성질 1.3.4로부터



례를 들어 이면 . 나머지 경우의 도 계산하여 표에 주었다.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | 1 | 2 | 3 | 4 | 5 | 6 | 합 |
|  | 0.0046 | 0.0324 | 0.0880 | 0.1713 | 0.2824 | 0.4213 | 1.0000 |

이것은 3개의 주사위를 던졌을 때 최대점수 는 우연량이며 가 6을 취할 확률은 0.4213이라는것을 보여준다. 또한 이므로 3개의 주사위를 던질 때 최대점수가 3점을 초과하지 않을 확률은 0.125밖에 안된다는것을 보여준다.

### 1.3.3 확률의 더하기공식

유한가법성이나 셀수 있는 가법성은 사건들이 서로 배반일 때 합사건의 확률을 구하는 공식을 준다. 그러면 일반적인 사건(반드시 서로 배반은 아니다)에 대해 어떻게 합사건의 확률을 구하겠는가? 다음 성질 1.3.6에서 식 (1.3.5)은 임의의 두개 사건의 합사건의 확률을 구하는 확률더하기공식을 주며 식 (1.3.6)은 임의의 개 사건의 합사건의 확률을 구하는 확률더하기공식을 준다. 이러한 성질들은 확률을 계산할 때 매우 쓸모있게 리용된다.

성질 1.3.6(더하기공식) 임의의 두 사건 , 에 대하여



임의의 개 사건 에 대하여



증명. 먼저 식 (1.3.5)을 증명하자. 이고 와 가 서로 배반이므로 유한가법성과 성질 1.3.5로부터 .

다음으로 귀납법으로 식 (1.3.6)을 증명한다. 일 때 식 (1.3.6)은 식 (1.3.5)이 된다. 식 (1.3.6)이 에 대하여 성립한다면 가정하자. 에 대하여 먼저 두 사건 와 에 식 (1.3.5)을 적용한다.



그리고 귀납가설로부터 와 에 대하여 전개하고 정돈하면 식 (1.3.6)이 에 대해서도 성립한다는것을 증명할수 있다.

따름(반가법성) 임의의 두 사건 에 대하여

.

임의의 개 사건 에 대하여

.

실례 1.3.4 알려진 사건 , , 의 확률이 각각 0.4, 0.3, 0.6이다. 를 구하시오.

풀기. 더하기공식 과 문제의 조건으로부터 . 따라서. 그러므로 식 (1.3.4)로부터 을 알수 있다.

실례 1.3.5 일 때 중에서 적어도 하나가 발생할 확률은 얼마인가? 또 가 모두 발생하지 않을 확률은 얼마인가?

풀기. 이고 또 이므로 확률의 단조성으로부터 이다. 다시 더하기공식을 적용하면 중 적어도 하나가 발생할 확률은



또한 "가 모두 발생하지 않는다"의 대립사건은 "가운데 적어도 하나는 발생한다"이므로 대립사건의 확률공식으로 얻을수 있다.

(가 모두 발생하지 않는다).

일반적으로 "적어도 하나는 발생한다"의 확률을 구할 때는 대립사건공식을 리용하는것이 편리하다. 그러나 아래의 실례 1.3.6의 정합문제는 대립사건을 리용하여 풀이를 구할수 없으며 "적어도 하나는 발생한다"는 사건을 사건들의 합으로 표시하고 일반사건의 더하기공식을 리용하여 풀이를 구해야 한다.

실례1.3.6(정합문제) 명이 참석한 저녁식사에 매 사람이 서로 다른 선물을 가져왔다고 가정하자. 저녁식사시간 매 사람이 한데 모아놓은 개의 선물중에서 임의로 하나를 선택한다. 적어도 한사람이 자기의 선물을 뽑을수 있는 확률은 얼마인가?

풀기. 를 사건 "()째 사람이 자기가 준비한 선물을 뽑는다"라고 하자. 구하려는 확률은 이다. 이므로



따라서 확률의 더하기공식 (1.3.6)으로부터 .

례컨대 일 때 이 확률은 0.63333이고 일 때 이 확률의 극한은 이다. 이것은 참가자가 많더라도(례를 들어 100명이상) 최소한 한명이 자기 선물을 뽑는 사건은 필연사건이 아니라는것을 의미한다.

### 1.3.4 확률의 련속성

확률의 련속성을 론의하기 위하여 먼저 사건렬의 극한에 대한 다음과 같은 정의를 준다.

정의 1.3.1

① 의 임의의 단조비감소하는 사건렬 에 대하여 셀수있는 합 을 의 **극한사건**이라고 하고 다음과 같이 표시한다.

.

② 의 임의의 단조비증가하는 사건렬 에 대하여 셀수 있는 사귐 을 의 **극한사건**이라고 하고 다음과 같이 표시한다.

.

이렇게 극한사건에 대한 정의를 한 다음 아래와 같이 확률함수의 련속성에 대한 정의를 내릴수 있다.

정의 1.3.2 우의 확률 에 대하여

① 만일 가 의 임의의 단조비감소하는 사건렬 에 대하여 모두 이 성립하면 확률 는 **하반련속**이라고 한다.

② 만일 가 의 임의의 단조비증가하는 사건렬 에 대하여 모두 이 성립하면 확률 는 **상반련속**이라고 한다.

그러면 이제 확률의 련속성을 증명할수 있다.

성질 1.3.7(확률의 련속성) 만일 가 사건령역 우의 확률이라면 는 하반련속이면서도 상반련속이다.

증명. 먼저 의 하반련속성을 증명하자. 를 의 단조비감소하는 사건렬이라고 하자. 즉 

라고 하면 . 이므로 들은 서로 둘씩 배반이다. 셀수 있는 가법성으로부터 . 따라서 . 이렇게 의 하반련속성이 증명되였다.

계속하여 의 상반련속성을 증명하자. 을 단조비증가하는 사건렬이라고 하면 은 단조비감소하는 사건렬이다. 따라서 확률의 하반련속성으로부터



따라서 이며 의 상반련속성이 증명된다.

아래에서 우리는 셀수 있는 가법성에 대하여 더 론의한다. 우로부터 알수 있듯이 셀수 있는 가법성으로부터 유한가법성과 하반련속성을 유도할수 있으나 유한가법성으로부터는 셀수 있는 가법성을 유도할수 없다. 이것은 유한가법성으로부터 셀수 있는 가법성을 유도하자면 일부 조건이 부족하다는것을 의미한다. 아래의 성질은 부족되는 조건이 바로 하반련속성이라는것을 설명하여준다.

성질 1.3.8 만일 가 우에서 을 만족하는 부가 아닌 모임함수라면 이 함수가 셀수 있는 가법성을 만족하기 위한 필요충분조건은 다음과 같다.

① 가 유한가법성을 만족한다.

② 가 하반련속이다.

증명. 성질 1.3.2와 성질 1.3.7로부터 필요성을 얻을수 있다. 충분성을 증명하여보자. 가 둘씩 서로 배반인 사건렬이라고 하면 유한가법성으로부터 임의의 유한한 에 대하여 이 성립한다. 이 같기식의 왼변은 1을 넘지 않는다.

따라서 정의 무한합렬 는 수렴한다. 즉

.

라고 표시하면 은 단조비감소하는 사건렬이고 하반련속성으로부터

.

식 (1.3.11)과 식 (1.3.12)을 결합하면 셀수 있는 가법성을 얻을수 있다.

성질 1.3.8에서 알수 있듯이 확률의 공리적정의에서 셀수 있는 가법성을 유한가법성과 하련속성으로 바꿀수 있다.

련습문제 1.3

1. 사건 와 가 서로 배반이고 일 때 다음과 같은 사건의 확률을 구하시오.

① 와 중 적어도 하나가 발생한다.

② 와 가 모두 발생한다.

③ 는 발생하지만 는 발생하지 않는다.

2. 이라고 가정하면 다음 명제가운데서 어느것이 정확한가?

① 와 는 배반이다.

② 와 는 배반이 아니다.

③ 는 불가능사건이다.

④ 는 반드시 불가능사건인것은 아니다.

⑤  또는 .

⑥ .

3. 제품을 1, 2, 3등급으로 나누는데 그가운데서 1등급제품은 2등급의 3배이고 3등급제품은 2등급의 절반이다. 이 제품가운데서 임의로 하나를 뽑을 때 3등급제품일 확률을 구하시오.

4. 0~9의 10개 수자가운데서 임의로 3개의 서로 다른 수자를 선택할 때 다음과 같은 사건의 확률을 구하시오.

① ={3개 수자가운데 0과 5가 포함되지 않는다}

② ={3개 수자가운데 0이나 5가 포함되지 않는다}

③ ={3개 수자가운데 0은 포함되지만 5는 포함되지 않는다}

5. 한 책방에서 A, B, C세 종류의 책을 팔고 있다. 구매자들가운데서 25%는 책 A, 20%는 책 B, 15%는 책 C, 10%는 책 A와 B, 8%는 책 A와 C, 5%는 책 B와 C, 3%는 책 A, B, C를 동시에 구매한다. 다음과 같은 사건의 확률을 구하시오.

① 책 A만 구매한다.

② 한 종류의 책만 구매한다.

③ 적어도 한 종류의 책을 구매한다.

④ 어떤 책도 구매하지 않는다.

6. 어떤 작업반에 남성 9명, 녀성 5명이 있다. 3명을 선출하려고 하는데 그 가운데 녀성이 적어도 1명 있을 확률은 얼마인가?

7. 주사위 한개를 4번 던져 적어도 한번 6점이 나타나는것과 주사위 두개를 24번 던져 적어도 한번 동시에 6점이 나타나는 기회는 동등하다고 하는 주장이 옳은가?

8. 수자 1~9에서 반복적으로 개의 수를 임의로 취할 때 그 수자들의 적이 10으로 완제될 확률을 구하시오.

9. 주머니속에 검은공 개와 흰공 1개가 있는데 매번 주머니에서 임의로 한개씩 뽑아내고 검은공 한개를 넣는다. 번째 만질 때 검은 공이 나올 확률은 얼마인가?

10. 이면 임의의 사건 에 대하여 라는것을 증명하시오.

11. 동전을 번 던져 앞면의 수가 뒤면의 수보다 많이 출현한 확률을 구하시오.

12. 세 사람이 있는데 매 사람이 모두 같은 확률 1/5로 5개 방의 임의의 한방에 배치받는다. 이때

① 세 사람이 모두 같은 방에 배치받을 확률을 구하시오.

② 세 사람이 서로 다른 방에 배정받을 확률을 구하시오.

13. 한 호실에 5명의 학생이 있는데 그들중 적어도 2명의 생일이 같은 달에 있을 확률을 구하시오.

14. 분대의 명 전사는 모두 개인이 보관리용하고있는 총을 1자루씩 가지고있으며 이 총들의 외형은 완전히 똑같다. 한차례의 야간비상소집에서 매 사람이 임의로 1자루씩 총을 가져올 때 적어도 1명이 자기의 총을 가져올 확률을 구하시오.

15. 두 사건 에 대하여 이다. 이때

① 어떤 조건하에서 가 최대값을 가지고 그 최대값은 얼마인가?

② 어떤 조건하에서 가 최소값을 가지고 그 최소값은 얼마인가?

16. 주어진 사건 가 을 만족시킨다. 라고 할 때 를 구하시오.

17. 일 때 를 구하시오.

18. 일 때 를 구하시오.

19. 임의의 사건 에 대하여 다음과 같은것을 증명하시오.

① . ② .

20. 세 사건 에 대하여 일 때 이라는것을 증명하시오.

21. 사건 의 확률이 모두 0.5이고 이면 이라는것을 증명하시오.

22. 다음의 관계식을 증명하시오.

① . ② .

23. 이라는것을 증명하시오.

## 1.4 조건부확률

조건부확률은 확률론에서 중요하고도 실용적인 개념이다.

### 1.4.1 조건부확률의 정의

조건부확률이란 어떤 사건 가 발생한 조건에서 다른 사건 가 발생할 확률로서 로 표시한다. 그것은 와 다른 종류의 확률로서 아래에서 실례를 들어 설명한다.

실례 1.4.1 어린이가 두명 있는 가정을 조사할 때 표본공간은 이다. 여기서 는 남자아이, 는 녀자아이, 는 첫 아이가 남자, 둘째 아이가 녀자라는것이며 기타 표본점도 류사하게 설명할수 있다.

에서 4개의 표본점이 등가능한 경우에 아래 사건들의 확률을 론의하기로 하자.

① 사건 ="집에 적어도 녀자아이가 한명 있다"의 발생확률은 이다.

② 만일 사건 ="집에 적어도 한명의 남자아이가 있다"가 발생하였다는것을 알고 다시 사건 가 발생할 확률을 구한다면 . 이것은 사건 의 발생으로 인하여 가 발생할 가능성이 배제되였기때문이며 이때 표본공간 도 로 변경되고 에서 사건 는 두개의 표본점을 포함한다. 따라서 . 이것이 바로 조건부확률이며 그것과 (무조건)확률 는 서로 다른 개념이다.

③ 우의 조건부확률에서 분자와 분모를 각각 4로 나누면 를 얻을수 있다. 여기서 사귐사건 는 "집에 남자아이 1명, 녀자아이 1명이 있다"이다.

이 관계를 일반화할수 있다. 즉 조건부확률은 2개의 무조건확률의 상인데 이것이 바로 조건부확률의 정의이다.

정의 1.4.1 와 가 표본공간 의 두 사건이고 이면

.

을 "**의 발생하에서 의 조건부확률**", 간단히 **조건부확률**이라고 부른다.

실례 1.4.2 표본공간 는 25개의 등가능한 표본점을 포함하고 사건 는 15개의 표본점을, 사건 는 7개의 표본점을, 사귐사건 는 5개의 표본점을 포함한다고 가정하자(그림 1.4.1). 이때 이다. 사건 가 발생한 조건에서 의 조건부확률은 .



*A*

*B*

*AB*

그림 1.4.1 실례 1.4.2의 벤도식

이 결과를 다음과 같이 고찰할수도 있다. 사건 가 발생했다는것은 사건 가 발생할수 없다는것을 의미하므로 의 18개 표본점은 무시될수 있다. 이때 의 7개 표본점가운데서 에 속하는것은 5개뿐이다. 따라서 . 이것은 조건부확률 가 표본공간 가 로 축소되였을 때 계산된다는것을 의미한다.

비슷하게 .

주의하여야 할 점은 조건부확률 는 가 주어진 조건에서 사건 의 확률을 론한다는것이다. 그렇다면 확률의 성질은 에 대하여 모두 성립되는가? 례를 들어



와 같은 확률의 성질들이 모두 성립되는가? 조건부확률이 3가지 공리를 만족시키는가만 검증할수 있으면 이 문제에 대답할수 있다.

성질 1.4.1조건부확률은 확률이다. 즉 이면 다음과 같다.

① .

② .

③ 의 들이 서로 배반이면 .

증명. 조건부확률의 정의로부터 ①과 ②를 쉽게 증명할수 있다. 아래에서 ③을 증명하겠다. 들이 서로 배반이므로 들도 서로 배반이다. 그러므로



곱하기공식, 완전확률공식, 베이스공식 등 조건부확률 특유의 세가지 쓸모있는 공식을 제시한다. 이 공식들은 복잡한 사건들의 확률을 계산하는데 도움을 줄수 있다.

### 1.4.2 곱하기공식

성질 1.4.2 곱하기공식

① 이면 다음과 같다.

.

② 이면



증명. 조건부확률의 정의로부터 정리하면 식 (1.4.2)를 얻을수 있다. 식 (1.4.3)을 증명하여보자. 이므로 식 (1.4.3)의 조건부확률이 모두 의의가 있고 조건부확률의 정의에 따라 식 (1.4.3)의 오른변은 다음과 같다.

.

따라서 식 (1.4.3)이 성립된다.

실례 1.4.3 한 조의 부속품이 총 100개이고 그중 10개가 불합격품이다. 그중에서 하나씩 꺼내는데 세번째만에 불합격품일 확률은 얼마인가?

풀기. 를 사건 "째로 꺼낸것이 불합격품이다"()라고 하면 구하려는 확률은 이고 곱하기공식으로부터 .

사실 실례 1.4.3은 아래의 실례 1.4.4의 특수경우이다.

실례 1.4.4(통모형) 통에 개의 검은 공과 개의 붉은 공이 있는데 매번 임의로 한개씩 꺼낸 후 그 공을 다시 넣고 같은 색의 공 개와 다른 색의 공 개를 넣는다. 를 사건 "번째로 꺼낸것이 검은 공이다", 를 사건 "번째로 꺼낸것이 붉은 공이다"라고 하자.

통에서 련속 3개의 공을 꺼낼 때 그중 2개는 붉은 공, 1개는 검은 공이라면 이 확률은 곱하기공식을 리용하여 얻을수 있다.







이 확률은 검은 공이 몇번째로 뽑히우는가와 관계가 있다.

통모형은 폴랴(pólya)모형으로도 불리우는데 이 모형을 다음과 같이 변화시킬수 있다.

①  즉 비반환표본화일 때 앞의 선택결과는 뒤의 선택결과에 영향을 주게 된다. 그러나 뽑아낸 검은 공과 붉은 공의 개수만 확정된다면 확률은 공을 뽑아낸 순서에 의존하지 않고 모두 같다. 이 실례에서 . 실례 1.4.3은 이 경우에 귀착된다.

②  즉 반환표본화일 때 앞의 선택결과는 뒤의 선택결과에 영향을 주지 않는다. 그러므로 우의 3개 확률은 모두 같다. .

③ 일 때 **전염병모형**이라고 한다. 이때 매번 공을 꺼낸 후 다음 번에 같은 색갈의 공을 얻을 확률이 증가된다. 다시 말하면 매번 전염병환자가 한명 발견될 때마다 그후에도 재전염될 확률이 증가된다. ①, ②와 마찬가지로 우의 세개 확률은 모두 같다.

.

우의 ①, ②, ③로부터 알수 있는바와 같이 통모형에서 이면 우의 세개 확률은 모두 같다. 즉 뽑아낸 검은 공과 붉은 공의 개수만 확정되면 확률은 공을 뽑아낸 순서에 의존하지 않고 모두 같다. 그러나 일 때에는 사정이 다르다.

④ 일 때 **안전모형**이라고 한다. 이 모형은 사고가 발생할 때(빨간 공을 꺼낼 때)마다 로동안전사업을 짜고들어 다음번 사고가 발생할 확률이 감소되고 반면에 사고가 발생하지 않았을 때(검은 공을 꺼냈을 때) 이 사업은 해이되여 다음번에 사고가 발생할 확률이 높아진다. 이 경우에 우의 세가지 확률은 각각 다음과 같다.



### 1.4.3 완전확률공식

완전확률공식은 확률론에서 중요한 공식의 하나로서 복잡한 사건의 확률을 간단하게 만들어 계산하는 효과적인 수단을 제공해준다.

성질 1.4.3(완전확률공식) 는 표본공간 의 분할(그림 1.4.2) 즉 이 서로 배반이고 이다. 만일 이면 임의의 사건 에 대해서도 다음의 관계식이 만족된다.

.

증명. 이고 들이 서로 배반이므로 가법성으로부터 이다. 다시 을 웃식에 대입하면 식 (1.4.4)가 얻어진다.

완전확률공식에 대하여 다음의 몇가지에 주의를 돌려야 한다.

① **완전확률공식의 가장 간단한 형식** 만일 이면 다음과 같다.

.

그림 1.4.3을 참고할수 있다.

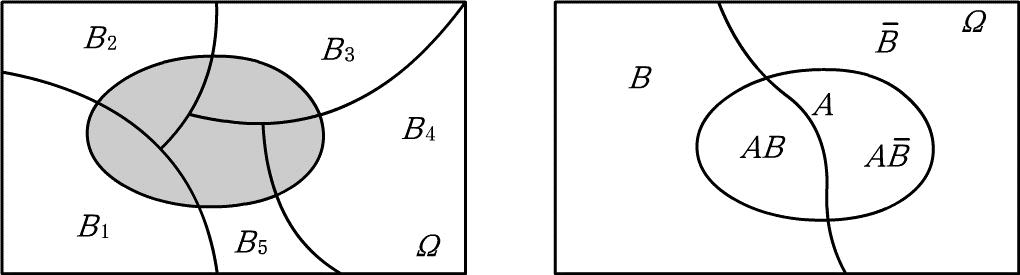


그림 1.4.2 표본공간의 한가지 분할()

그림 1.4.3 와 를 리용한 표본공간의 분할

② 가 표본공간의 분할이라는 조건을 들이 서로 배반이고 이라는것으로 바꾸어도 완전확률공식은 성립한다.

③ 셀수 있는 사건 들이 서로 배반이고 이면 완전확률공식은 여전히 성립하며 이때 식 (1.4.4)의 오른변을 셀수 있는 항들의 합으로 고쳐쓰기만 하면 된다.

실례 1.4.5(추첨모형) 장의 복권가운데 한장이 당첨된다. 두번째 사람이 당첨복권을 뽑을 확률이 얼마인지 구하시오.

풀기. 를 사건 "번째 사람이 당첨복권을 뽑는다"()이라고 하면 목적은 을 구하는데 있다. 의 발생여부는 의 발생확률과 직접 관계되기때문에 . 또한 과 은 모두 확률이 0보다 큰 사건들이다. . 그리하여 완전확률공식으로부터

.

이것은 복권에 당첨될 기회는 순서와 무관계하다는것을 의미한다. 왜냐하면 후에 뽑는 사람은 "불리한 상황"(앞 사람이 이미 당첨복권을 뽑았다)에 처해있을수도 있고 "유리한 상황"(앞 사람이 아직 당첨복권을 뽑지 않았으므로 뒤 사람이 당첨복권을 뽑을 기회가 증가한다)에 처해있을수도 있는데 두 상황을 완전확률공식으로 종합(무게평균)하여 얻은 결과는 공평하고 합리적이기때문이다.

비슷한 방법으로 이라는것을 알수 있다.

만일 장의 복권가운데 장이 당첨될수 있다면 .

이것은 복권을 구매할 때 먼저 구매하든 후에 구매하든 당첨기회는 균등하다는것을 말해준다.

실례 1.4.6 보험회사는 어떤 보험종류의 보험계약자를 사고를 일으키기 쉬운 보험자와 안전한 보험자의 두가지로 분류할수 있다고 본다. 통계적인 결과에 의하면 사고가 잦은 사람이 1년안에 사고가 발생할 확률은 0.4 이지만 안전한 사람은 0.1밖에 안된다. 만일 사고를 일으키기 쉬운 보험자가 보험계약자들중에서 차지하는 비률이 20%라고 가정하면 현재 새로운 보험계약자가 이 보험에 가입할 경우 그 보험계약자가 보험계약서를 구매한후 1년내에 사고가 발생할 확률이 얼마나 되는가?

풀기. ="보험계약자가 1년내에 사고를 낸다", ="보험계약자가 사고를 일으키기 쉬운 사람이다"라고 하면 ="보험계약자가 안전한 사람이다"이며 이다. 완전확률공식으로부터 이다.

실례 1.4.7(민감성문제조사) 학생이 시험과정에 부정행위를 하는 행위는 모두 교원과 동무들을 피해 진행하는것으로서 개인적인 사생활에 속한다. 현재 조사방안을 설계하여 조사자료로부터 학생들이 부정행위를 한 비률 를 추정해야 한다. 민감성문제에 대한 조사방안에서 중요한것은 피조사자가 진실하게 대답하고 또 개인비밀을 지키도록 하는것이다. 조사방안이 잘못 설계되면 조사대상자는 협조를 거부하고 조사수치의 진실성을 잃게 된다. 다년간의 연구와 실천을 거쳐 일부 심리학자와 통계학자들은 한가지 조사방안을 설계하였는데 응답자들은 아래의 두가지 질문중 하나만 대답하면 되며 또 "예" 또는 "아니"라고 대답하면 된다. (소아마비)

질문 A: 당신의 생일이 7월 1일 이전인가?

질문 B: 부정행위를 한 적이 있는가?

이 조사방안은 간단한것 같지만 피조사자의 우려를 없애고 피조사자가 이번 조사에서 개인비밀이 루설되지 않을것이라는 확신을 가지도록 하기 위하여 다음과 같은 조작을 할수 있다.

① 조사대상자가 옆에 사람이 없는 빈방에서 혼자 조작하고 질문에 대답하게 한다.

② 조사대상자가 흰 공과 빨간 공만 있는 통에서 임의로 공 한개를 뽑아 색을 보고 다시 넣는다.

흰 공을 뽑았으면 문제A, 빨간 공을 뽑았으면 문제 B에 대답하도록 한다.

조사대상자가 문제 A에 대답하든 문제 B에 대답하든 그림 1.4.4와 같은 답안만 있으면 된다. 답안의 네모칸에 표식을 한후 그것을 밀봉한 투표통에 넣도록 한다. 이런 방법은 대답한 문제가 A인지 B인지 주변 사람들이 알수 없게 하여 피조사자의 불안감을 해소시킬수 있다.

예

아니

답안지

그림 1.4.4 민감성문제의 답안지

이제 조사결과를 분석하여보자. 물론 문제 A에 대하여 흥미가 없다.

우선 장의 답안지(이 비교적 크며 례컨대 1000장이상)를 가지고있으며 그중 장은 "예"라고 대답하였다고 하자. 이 장의 답안지가운데서 몇장이 문제 B에 대답했는지, 또 장의 "예"답안지가운데서 몇장이 문제 B에 대답했는지 알수 없다. 하지만 두가지 정보를 미리 알고 있다. 즉

① 참가인원수가 비교적 많은 경우에 임의의 한사람의 생일이 7월 1일 이전일 확률은 0.5이다.

② 통의 붉은 공의 비률 은 이미 알고있다.

이제 이 4개의 자료 를 리용하여 를 구해야 한다. 완전확률공식으로부터 (예)=(흰 공)(예 | 흰 공) + (빨간 공)(예 | 빨간 공)을 얻는다. 그러므로 (빨간 공)=, (흰 공)=, (예 | 흰 공)=0.5, (예 | 빨간 공)=를 웃식에 갈아넣고 오른변을 잦음률 으로 바꾸면 . 이로부터 .

확률 (예)를 잦음률 로 교체하였기때문에 웃식으로부터 에 대한 추정값을 얻을수 있다.

례를 들어 실지 조사에서 통안에 빨간 공 30개, 흰 공 20개를 넣었을 때 이며 조사가 끝난 후 총 1583장의 답안지가운데서 389장이 "예"답안지라면 .

### 1.4.4 베이스공식

곱하기공식과 완전확률공식으로부터 다음과 같은 유명한 공식을 유도할수 있다.

성질 1.4.4(베이스공식) 가 표본공간 의 분할 즉 가 서로 배반이고 이라면 일 때

.

증명. 조건부확률의 정의로부터 . 이 식의 분자에 대하여 곱하기공식을, 분모에 대하여 완전확률공식을 적용하면 . 따라서 이 얻어진다.

실례 1.4.8 어느 지역 주민의 간암발병률은 0.0004이다. 화학적검사에 의하여 조사를 진행할 때 오차가 있을수 있다. 간암이 있는 사람의 화학적검사결과는 99%가 양성이고 간암이 없는 사람의 검사결과는 99.9%가 음성이다. 어떤 사람의 검사결과가 양성일 경우 진짜로 간암에 걸렸을 확률은 얼마인가?

풀기. 를 사건 "피검사자에게 간암이 있다", 를 사건 "검사결과가 양성이다"라고 하자. 이때



목적이 를 구하는것이므로



이것은 검사결과가 양성인 사람들가운데서 진짜 간암이 있는 사람은 30%도 안된다는것을 의미한다. 놀랄만한 결과이지만 자세히 생각해보면 리해가 된다. 간암에 걸리는 사람은 인구 1만명가운데 4명정도로서 약 9996명은 간암에 걸리지 않을 정도로 발병률이 낮기때문이다. 10000명에 대하여 화학적검사를 진행하였는데 오검사확률에 따라 간암이 없는 9996명중 약 9996×0.001=9.996명이 양성반응을 보였다. 또한 4명의 진짜 간암 환자중 약 4×0.99=3.96명은 양성으로 보고되였다. 13.956명의 양성환자가운데서 진짜 간암환자는 약 3.96명으로서 28.4%를 차지한다.

오검사확률을 더 낮추는것은 검사정확도를 높이는데서 중요하다. 실제로 기술과 조작 등 여러가지 원인으로 하여 오검사확률을 낮추는것은 매우 어렵다. 때문에 흔히 재검사의 방법으로 오유률을 감소시킨다. 또는 다른 간단하고 시행하기 쉬운 보조방법으로 먼저 초보적인 검사를 진행하여 간암이 아닌 많은 사람들을 배제한후 다시 화학적검사법으로 의심되는 대상에 대해 검사를 진행할수 있다. 이때 의심대상들의 모임에서 간암발병률은 크게 높아진다. 례를 들어 1차 양성판정을 받은 사람을 대상으로 재검사를 할 때 인데 베이스공식을 리용하면 이다. 이렇게 화학적검사법의 정확도가 크게 향상될수 있다.

우의 례에서 사건 를 "원인"으로 보고 사건 를 "결과"로 본다면 베이스공식을 리용하여 "결과"를 알고 있는 조건하에서 "원인"의 확률 를 구한것으로 된다. "결과"의 (무조건)확률 를 구하려면 완전확률공식을 리용한다. 우의 례에서 를 취하면 다음과 같다.



조건부확률의 3개 공식가운데서 곱하기공식은 사귐사건의 확률을 구하는것이고 완전확률공식은 복잡한 사건의 확률을 구하는것이며 베이스공식은 조건부확률을 구하는것이다.

베이스공식에서 를 의 **사전확률**이라고 하고 를 의 **사후확률**이라고 한다면 베이스공식은 전문적으로 사후확률을 계산하는데 리용되는것으로 볼수 있는데 의 발생이라는 새로운 정보를 통하여 의 확률에 수정을 진행한다. 다음의 실례는 이것을 보여주고있다.

실례 1.4.9 이소프의 우화 "아이와 승냥이"에서 한 아이가 매일 산으로 양을 방목하러 가는데 산에 승냥이가 나타나군 한다. 어느날 그가 산우에서 "승냥이가 온다!"고 소리를 쳐서 산밑에서 일하던 농군들이 일손을 놓고 승냥이를 잡으러 갔는데 산에 올라가니 승냥이가 없었다. 이튿날에도 역시 그러하였다. 사흘째 되는 날 승냥이가 정말로 왔다. 그러나 아이가 아무리 소리쳐도 누구도 그를 구하러 오지 않았다. 그것은 두번씩이나 그가 거짓말을 했기때문에 사람들은 그를 더 이상 믿지 않았기때문이다.

우화에서 이 아이에 대한 마을사람들의 신뢰도가 어떻게 떨어졌는가를 베이스공식으로 분석해보자.

우선 사건 를 "아이가 거짓말을 한다", 사건 를 "아이가 믿을만하다"라고 하자. 마을사람이 과거에 이 아이에 대해 어떤 인상을 가지고있었다고 가정하자. 즉

.

베이스공식을 리용하여  즉 아이가 한번 거짓말을 한후에 사람들이 그에 대한 신뢰의 변화를 보자.

베이스공식에서 와 를 리용하는데 는 믿을만한 아이()가 거짓말()할 가능성이고 는 믿을수 없는 아이()가 거짓말()할 가능성이다. 편리하게 라고 하자.

처음으로 사람들이 산에 올라가 승냥이를 잡으려 했는데 승냥이가 오지 않자 아이가 거짓말을 했다()고 여기고 이 정보에 근거하여 아이에 대한 신뢰도를 베이스공식을 리용하여 갱신한다.

.

즉 한번 속고 난 다음 아이에 대한 믿음이 0.8로부터 0.444로 된다. 즉 식 (1.4.7)로부터

.

이 기초우에서 베이스공식을 다시 적용하여 를 계산하는데 이것은 아이가 두번 거짓말을 하였을 때 사람들의 신뢰도가 다시 갱신된다는것을 의미한다.

.

이것은 사람들이 두번 속으면서 아이에 대한 신뢰도가 0.8로부터 0.138로 떨어졌음을 의미한다. 이렇게 신뢰도가 낮은데 어떻게 세번째 호출을 받고 산에 올라가겠는가.

련습문제 1.4

1. 어느한 학급 학생들의 시험성적에서 수학이 불합격인 학생이 8%이고 어문이 불합격인 학생이 5%이며 이 두 과목에서 모두 불합격인 학생이 2%이다.

① 수학에서 불합격된 학생이 어문에서도 불합격될 확률은 얼마인가?

② 어문에서 불합격된 학생이 수학에서도 불합격될 확률은 얼마인가?

2. 1, 2, 3등급의 제품이 각각 60%, 35%, 5%를 차지한다고 하자. 그중에서 임의로 하나를 꺼냈는데 3등급의 제품은 아니다. 그것이 1등급 제품일 확률을 구하시오.

3. 두개의 주사위를 던지는데 를 사건 "두 점수의 합이 10이다", 사건 를 "첫알의 점수가 두번째 알의 점수보다 작다"라고 하면 조건부확률 를 구하시오.

4. 어떤 동물이 태여나서 10년까지 살 확률을 0.8이라고 하고 15년까지 살 확률을 0.5라고 하자. 지금 10년생인 이 동물이 15년까지 살 확률을 구하시오.

5. 10개의 제품가운데 3개가 불합격품이라고 할 때 두개를 임의로 선택하였다. 그중 하나가 불합격품이라는것을 알고 다른 하나도 불합격품일 확률을 구하시오.

6. 개의 제품가운데 개의 불합격품이 있다. 그중 두개를 임의로 취할 때 한개가 합격품이라는것을 알고 다른 한개도 합격품일 확률을 구하시오.

7. 한 주사위를 두번 던졌을 때 가 각각 차례로 던진 점수라고 하고 라고 하자. 를 구하시오.

8. 일 때 를 구하시오.

9. 일 때 를 구하시오.

10. 두 사건 에 대하여 일 때 를 구하시오.

11. 주머니속에 흰 공 1개와 검은 공 1개가 있다. 그중 한개를 임의로 선택하는데 만일 흰 공을 꺼내면 실험은 중지된다. 검은 공을 꺼냈을 경우 꺼낸 검은 공을 다시 넣고 또 하나의 검은 공을 추가한다. 흰색이 나올 때까지 반복할 때 아래와 같은 사건의 확률을 구하시오.

① 번째까지 꺼냈는데 실험이 끝나지 않는다.

② 번째를 꺼냈을 때 실험이 끝난다.

12. 전자관 한 상자에 합격품이 8개, 불합격품이 2개 있다. 그중에서 반환하지 않고 하나씩 꺼낼 때 두번째로 꺼낸 제품이 합격품일 확률을 구하시오.

13. 주머니 A에는 흰 공 개와 검은 공 개가 있고 주머니 B에는 흰 공 개와 검은 공 개가 있다.

① 주머니 A에서 임의로 공 하나를 꺼내 주머니 B에 넣고 또 주머니 B에서 임의로 공 하나를 꺼낸다. 마지막에 주머니 B에서 꺼낸 공이 흰 공일 확률을 구하시오.

② 주머니 A에서 공 2개를 임의로 꺼내 주머니 B에 넣고 또 주머니 B에서 임의로 공 하나를 꺼낸다. 마지막에 주머니 B에서 꺼낸 공이 흰 공일 확률을 구하시오.

14. 개의 주머니가 있는데 매 주머니마다 흰 공 개와 검은 공 개가 있다. 첫 주머니에서 임의로 공을 한개 꺼내여 두번째 주머니에 넣고 두번째 주머니에서 다시 임의로 공을 한개 꺼내 세번째 주머니에 넣는 식으로 해나가면 마지막에 번째 주머니에서 임의로 꺼낸 공이 흰 공일 확률을 구하시오.

15. 열쇠를 떨어뜨릴 확률이 기숙사에는 50%, 교실에는 30%, 길에는 20%이다. 그러나 세 곳에서 열쇠를 찾을 확률은 각각 0.8, 0.3, 0.1이다. 열쇠를 찾을 확률을 구하시오.

16. 두대의 선반이 같은 부속품을 가공할 경우 첫 선반과 두번째 선반에서 불합격품이 나타날 확률은 각각 0.03, 0.06이다. 가공한 부속품들을 한데 모아 놓았는데 첫 선반으로 가공한 부속품이 두번째 선반으로 가공한 부속품의 2배라는것을 알고 있다.

① 임의로 선택한 부속품이 합격품일 확률을 구하시오.

② 꺼낸 부속품이 불합격품이면 그것이 두번째 선반에서 가공된것일 확률을 구하시오.

17. 부속품이 두 상자 있는데 첫 상자에는 50개 있으며 그중 20개는 1등품이다. 두번째 상자에는 30개 있는데 그중 18개가 1등품이다. 이제 두 상자가운데서 임의로 한 상자를 선택하고 그 상자에서 차례로 두개의 부속품을 꺼내 시험한다.

① 처음 꺼내는 부속품이 1등품일 확률을 구하시오.

② 처음 꺼낸것이 1등급인 조건하에서 두번째 꺼낸 부속품이 또 1등급일 확률을 구하시오.

18. 학생이 4개 선택항목이 있는 단항선택문제를 풀 때 만일 문제의 정확한 답안을 모른다면 우연적으로 선택한다. 이제 시험지를 보고 다음의 상황에서 학생이 정확한 답안을 확실히 알고있을 확률을 구하시오.

① 학생이 정확한 답을 알거나 마음대로 선택한 확률이 모두 1/2이다.

② 학생이 정확한 답을 알 확률은 0.2이다.

19. 남자들가운데 5%, 녀자들가운데 0.25%가 색맹이라고 한다. 남녀비률이 22:21인 사람들가운데서 임의로 한명을 뽑았는데 색맹환자라고 한다. 이 사람이 남자일 확률을 구하시오.

20. 주머니속에 검은색인지 흰색인지 알수 없는 공이 하나 있다. 이제 주머니속에 흰 공 하나를 넣은 다음 주머니속에서 임의로 하나를 꺼냈는데 그것이 흰 공이였다. 주머니속에 있던 공이 흰 공일 가능성은 얼마나 되는가?

21. 개 바줄의 2개의 끝을 임의의 두개씩 련결할 때 정확히 개의 원을 이룰 확률을 구하시오.

22. 명의 사람들이 서로 공을 주고받는다. A의 손으로부터 공이 넘어가며 매번 넘길 때 다른 명의 사람들중의 어느 한 사람에게 등가능하게 공을 넘겨준다. 번째 넘기기를 할 때 A로부터 공이 넘어갈 확률을 구하시오.

23. 두 사람 A, B가 번갈아 주사위를 던지는데 A가 먼저 던진다. 어떤 사람이 주사위를 던져 1점이 나오면 상대방이 던지고 그렇지 않으면 그 사람이 계속 던진다. 번째 주사위를 A가 던질 확률을 구하시오.

24. 주머니 A에는 검은 공 1개와 흰 공 2개가 있고 주머니 B에는 흰 공 3개가 있다. 매번 두개의 주머니에서 각각 공을 임의로 한개씩 꺼내서 교체한 다음 다른 주머니에 넣는다. 번 교체한 다음 검은 구가 그냥 주머니 A에 있을 확률을 구하시오.

25. 두가지 날씨상황(비가 온다는것과 오지 않는다는것)만 고려한다고 가정하자. 현재의 날씨를 안다면 다음 날의 날씨가 변하지 않을 확률은 이고 변할 확률은 이다. 첫날에 비가 오지 않는다고 가정하고 번째 날에도 비가 오지 않을 확률을 구하시오.

26. 통에 개의 검은 공과 개의 빨간 공이 있다. 매번 임의로 한개씩 꺼내보고 꺼냈던 원래 공을 다시 넣으며 또 개의 같은 색 공을 넣는다. 번째로 꺼낸 공이 검은 공일 확률이 이라는것을 증명하시오.

27. 주머니속에 흰 공 개, 검은 공 개, 빨간 공 개가 있는데 주머니로부터 하나씩 반환하지 않고 공을 꺼낸다. 흰색이 검은색보다 빨리 나타날 확률은 이며 과는 무관하다는것을 증명하시오.

28. 라면 이라는것을 증명하시오.

29. 사건 와 가 서로 배반이고 이면 이라는것을 증명하시오.

30. 를 임의의 두 사건이라고 할 때 이면 임을 증명하시오.

31. 이면 임을 증명하시오.

32. 라고 할 때 이라는것을 증명하시오.

33. 이면 이라는것을 증명하시오.

## 1.5 독립성

독립성은 확률론의 또 하나의 중요한 개념으로서 독립성을 리용하여 확률계산을 간소화할수 있다. 먼저 두 사건사이의 독립성에 대해 론의하고 이어서 여러 사건들사이의 호상독립성을 론의하며 마지막으로 실험들사이의 독립성에 대해 론의한다.

### 1.5.1 두 사건의 독립성

두 사건사이의 독립성이란 한 사건의 발생이 다른 한 사건의 발생에 영향을 주지 않는것을 말한다. 이것은 현실문제에서 매우 많은데 례를 들어 두개의 주사위를 던지는 실험에서 사건 를 "첫 주사위의 점수가 1이다", 사건 를 "둘째 주사위의 점수가 4이다"라고 하면 분명히 와 의 발생은 호상 영향을 주지 않는다.

확률의 관점에서 볼 때 사건 의 조건부확률 와 무조건확률 의 차이점은 사건 의 발생이 사건 가 발생하는 확률을 변화시킨다는데 있다. 즉 사건 는 사건 에 일정한 "영향"을 준다. 만일 사건 와 의 발생이 서로 영향을 미치지 않는다면 이며 이것은 다음의 관계식과 동등하다.

.

또한  혹은 인 경우에도 식 (1.5.1)은 성립한다. 따라서 식 (1.5.1)을 두 사건의 독립성에 대한 정의로 리용한다.

정의 1.5.1 식 (1.5.1)이 성립하면 사건 와 는 **서로 독립**이라고 하며 간단히 와 는 **독립**이라고 한다. 그렇지 않을 경우 와 는 **비독립**이라고 하거나 **서로 의존**한다고 한다.

많은 현실문제에서 두 사건이 서로 독립이라는것은 대부분 경험(서로 영향을 주고받는 여부)에 근거하여 판단한다. 례를 들면 우의 두개 주사위에 대한 문제에서 와 의 독립성이다. 그러나 어떤 경우에는 식 (1.5.1)을 리용하여 이 두 사건사이의 독립성을 판단하기도 한다.

실례 1.5.1 독립인 사건들의 례

① 52장의 주패 한조에서 임의로 한장을 뽑을 때 를 사건 "스페이드를 뽑는다", 를 사건 "J를 뽑는다"라고 하자. 이므로 "스페이드 J를 뽑는다"에 대한 사건 의 확률은 이고 와 는 서로 독립이다.

② 아이가 셋인 가정들을 고려하여 8가지 가능한 경우 는 모두 등가능하다. 그중에서 는 남자아이, 는 녀자아이를 나타낸다. 를 사건 "집안에 남자아이와 녀자아이가 모두 있다", 를 사건 "집안에 녀자아이가 많아서 한명 있다"라고 하면 이므로 "집안에 외동딸이 있다"에 대한 사건 의 확률은 이고 따라서 와 는 서로 독립이다.

③ 아이가 두명인 가정들을 조사한다면 표본공간은 4개의 표본점 만을 포함한다. 만일 사건 , 가 ②에서와 같은 경우 이고 이다. 이므로 와 는 독립이 아니다.

성질 1.5.1 사건 와 가 독립이면 사건 와 , 와 , 와 가 독립이다.

증명. 확률의 성질로부터 이고 또 와 의 독립성으로부터 이므로 . 이것은 와 가 독립이라는것을 말해준다. 류사하게 와 , 와 가 서로 독립이라는것을 증명할수 있다.

성질 1.5.1은 직관적으로 리해하기 쉽다. 와 가 서로 독립이기때문에 의 발생은 의 발생에 영향을 미치지 않으며 따라서 의 발생은 가 발생하지 않는데 영향을 미치지 않고 가 발생하지 않는것도 의 발생에 영향을 미치지 않으며 가 발생하지 않는것이 가 발생하지 않는데 영향을 미치지 않는다.

### 1.5.2 여러 사건의 호상독립성

우선 세 사건의 호상독립성에 대해 연구하며 이에 대하여 먼저 다음과 같은 정의를 내린다.

정의 1.5.2 를 세 사건이라고 하고



이 성립하면 는 **둘씩 서로 독립**이라고 하며

.

이면 는 **호상독립**이라고 한다.

주의: 반례를 가지고 식 (1.5.2)으로부터 식 (1.5.3)을 유도할수 없고 마찬가지로 식 (1.5.3)으로부터 식 (1.5.2)을 유도할수 없다는것을 증명할수 있다.

이에 따라 세개 이상 사건들의 호상독립성을 정의할수 있다.

정의 1.5.3 개의 사건 들에 대하여 임의의 에 대하여 다음과 같은 같기식이 만족된다면 개의 사건 들이 호상독립이라고 한다.



우의 정의에서 볼수 있다싶이 개의 서로 독립인 사건들의 임의의 부분안에서 사건들이 역시 서로 독립일뿐만아니라 임의의 부분과 다른 부분도 서로 독립이다. 성질 1.5.1과 류사하게 서로 독립된 사건중에서 임의의 부분을 대립사건으로 바꾸어도 얻은 사건들은 여전히 서로 독립이라는것을 증명할수 있다.

실례1.5.2 세 사건 이 서로 독립이라고 할 때 와 가 서로 독립이라는것을 증명하시오.

증명.



이므로 와 는 서로 독립이다.

이 문제의 증명과 류사하게 와 가 독립이고 와 가 독립이라는것을 쉽게 유도할수 있다. 주의할것은 만일 사이에 둘씩만 독립이라면 와 , 와 , 와 가 서로 독립이라는것을 증명할수 없다는것이다.

실례 1.5.3 두 사격수가 서로 독립적으로 동일한 목표에 사격을 한다. ㄱ의 명중확률을 0.9, ㄴ의 명중확률을 0.8이라고 하면 목표가 명중될 확률은 얼마인가?

풀기. 를 사건 "ㄱ가 목표를 명중한다"로, 를 사건 "ㄴ가 목표를 명중한다"로 표시하면 사건 "목표가 명중된다"는 이고 따라서 .

이 문제를 대립사건공식으로도 풀수 있는데 구체적으로 다음과 같다.



실례 1.5.4 두가지 공정으로 부속품을 가공한다. 첫번째 공정에는 3개의 처리가 있으며 매 처리에서 불합격품이 나타날 확률은 각각 0.3, 0.2, 0.1이다. 두번째 공정에서는 두개의 처리를 거치는데 불량품이 발생하는 발생할 확률은 각각 0.3, 0.2이다. 이때

① 어떤 공정으로 합격품을 가공해낼 확률이 더 큰가?

② 두번째 공정의 두 처리에서 불합격품이 나타날 확률이 모두 0.3일 경우 상황은 어떠한가?

풀기. 를 사건 "째 공정으로 합격품을 가공해낸다"()라고 하자.

① 매 공정이 독립적으로 작업한다고 볼수 있기때문에



즉 두번째 공정에서 합격품을 얻을 확률이 더 크다. 이 결과도 리해할만하다. 왜냐하면 두번째 공정의 앞의 두 처리에서 불합격품이 나타날 확률은 첫번째 공정과 같지만 한 처리가 없기때문에 불합격품이 생길 기회가 줄어들었기때문이다.

② 두번째 공정의 두 처리에서 불합격품이 나타날 확률이 모두 0.3일 때 이므로 첫번째 공정으로 합격품을 얻을 확률이 더 크다.

실례 1.5.5 두 선수가 번갈아 동일한 목표를 향해 사격하는 경우 ㄱ, ㄴ가 목표를 명중시킬 확률은 각각 , 이다. ㄱ가 먼저 쏘며 먼저 명중시키는 선수가 이긴다. ㄱ와 ㄴ가 이길 확률은 각각 얼마인가?

풀기 1. 를 사건 "째 사격으로 목표를 명중한다"()라고 하면 ㄱ가 먼저 사격하기때문에 사건 ="ㄱ가 이긴다"는 로 표시할수 있고 또 매 사격이 서로 독립이므로



마찬가지로 사건 ="ㄴ가 이긴다"에 대하여



이 문제를 계산할 때 조건 이 있어야 하는데 이것은 현실에서 쉽게 만족된다. 왜냐하면 가 0이나 1을 취하는것은 의미가 없기때문이다.

풀기 2. 이므로 이다. 그리하여

.

실례 1.5.6 체계는 여러 개의 블로크들로 구성되여있으며 모든 블로크는 독립적으로 동작한다. 매 블로크가 정상동작하는 확률을 라고 하고 다음의 체계들이 정상동작하는 확률을 구하시오.

① 직렬련결체계 

② 병렬련결체계 

③ 5개 블로크로 구성된 다리식체계 

풀기. ="번째 체계가 정상동작한다", ="번째 블로크가 정상동작한다"라고 하자.

1

2



1

2



1

2

4

5

3



① 직렬체계의 경우 "체계가 정상적으로 동작한다"는것은 "모든 블로크들이 정상동작한다"는것과 같고 그래서 이며 이다.

이것으로부터 정상동작확률이 0.9인 2개의 블로크로 구성된 직렬체계는 정상동작할 확률이 0.81로 떨어진다는것을 알수 있다.

② 병렬체계의 경우 "체계가 정상적으로 동작한다"는것은 "적어도 하나의 블로크가 정상동작한다"는것과 같고 그래서 이며 이다.

또는 

이것으로부터 정상동작확률이 0.9인 2개의 블로크로 구성된 병렬체계은 정상동작하는 확률이0.99로 제고된다는것을 알수 있다.

③ 다리식체계에서는 세번째 블로크가 중요한데 먼저 완전확률공식을 리용하면

.

세번째 블로크가 정상동작한다는 조건밑에서 체계는 선병렬후직렬체계(그림 1.5.1)가 되므로



또한 세번째 블로크가 정상동작하지 않는다는 조건밑에서 체계는 선병렬후직렬체계(그림 1.5.2)가 되므로 .

결국 

1

2

4

5

1

2

4

5

그림 1.5.1 선병렬후직렬체계

그림 1.5.2 선직렬후병렬체계

### 1.5.3 실험의 독립성

사건의 독립성을 리용하여 두개 이상의 실험들사이에 독립성을 정의할수 있다.

정의 1.5.4 두개의 실험 이 있다. 만일 실험 의 임의의 결과(사건)와 실험 의 임의의 결과(사건)이 모두 서로 독립인 사건이라면 이 **두 실험은 서로 독립**이라고 한다.

례를 들면 한개의 동전을 던지는것(실험 )과 한개의 주사위를 던지는것(실험 )은 서로 독립인 실험이다.

이와 비슷하게 개의 실험 의 호상독립성도 정의할수 있다. 만일 의 임의의 결과, 의 임의의 결과, ……, 의 임의의 결과들이 모두 서로 독립인 사건들이면 실험 는 호상독립이라고 한다. 만일이 개의 독립인 실험들이 같은것들이면 이것을 **중독립중복실험**이라고 한다. 만일 **중독립중복실험**에서 매 실험의 가능한 결과가 , 의 두가지이면 이 실험을 **중베르누이시행**(Bernoulli experiment)이라고 한다.

례를 들어 개의 동전을 던지는것, 개의 주사위를 던지는것, 개의 제품을 검사하는것 등은 모두 중독립중복실험이다.

실례 1.5.7 어떤 복권을 매주 1회 추첨하며 매번 당첨가능성은 10만분의 1, 매 주의 추첨은 서로 독립이다. 만일 매주 한장씩 10년(52주)동안 복권을 산다면 한번도 당첨되지 못할 가능성은 얼마나 되는가?

풀기. 매번 당첨될 가능성은 이므로 매번 당첨되지 못할 가능성은 이다. 또한 10년동안 모두 520번 복권을 구매하며 매번 추첨은 서로 독립이여서 520차의 독립중복실험을 진행한것과 같다. 를 "번째 추첨에서 당첨되지 못한다"()이라고 하면 은 서로 독립이고 따라서 10년동안 한번도 당첨되지 못할 가능성은 .

이것은 10년동안 복권에 당첨되지 못하는것이 례상사라는것을 의미한다.

만일 매번 당첨기회가 만분의 1이라면 10년동안 한번도 당첨되지 못할 가능성은 0.9493로서 크다.

련습문제 1.5

1. 3명이 독립적으로 하나의 암호를 해득한다. 그들이 단독으로 해득해낼수 있는 확률은 각각 1/5, 1/4, 1/3이다. 이 암호를 해득해낼 확률을 구하시오.

2. 두 종류의 종자가 있고 발아률은 각각 0.8과 0.9이다. 두 종류의 종자에서 각각 한알씩 임의로 선택할 때 다음과 같은 확률을 구하시오.

① 두알의 종자가 모두 발아할수 있는 확률.

② 적어도 한알의 종자가 발아할수 있는 확률.

③ 꼭 한알의 종자가 발아할수 있는 확률.

3. A와 B가 독립적으로 동일한 목표물을 한번 사격할 때 명중률은 각각 0.8, 0.7이다. 목표가 이미 명중되였다는것을 알고 그것이 A가 명중하였을 확률을 구하시오.

4. 회로가 세 소자 로 구성되였고 매 소자 에서 고장이 발생하는 확률은 각각 0.3, 0.2, 0.2이다. 그리고 매 소자들이 독립적으로 동작한다면 다음과 같은 경우에 이 회로에서 고장이 발생할 확률을 구하시오.

① 3개 소자 를 직렬련결한다.

② 3개 소자 를 병렬련결한다.

③ 소자 와 2개의 병렬련결된 소자 를 직렬련결한다.

5. 한주일내에 3대의 공작기계 가 수리해야 할 확률은 각각 0.9, 0.8, 0.85이다. 이때 한주일내에

① 수리할 공작기계가 하나도 없을 확률을 구하시오.

② 적어도 한대의 공작기계를 수리하지 않아도 되는 확률을 구하시오.

③ 기껏 한대의 공작기계를 수리해야 할 확률을 구하시오.

6. 이 서로 독립이고 일 때 가운데서

① 적어도 하나가 나타날 확률을 구하시오.

② 꼭 하나가 나타날 확률을 구하시오.

③ 기껏 하나가 나타날 확률을 구하시오.

7. 사건 와 가 서로 독립이고 배반이라면 를 구하시오.

8. 일 때 다음과 같은 경우에 를 구하시오.

① 와 가 서로 배반이다.

② 와 는 서로 독립이다.

③ .

9. 가 둘씩 독립이고 이다. 이때

① 일 경우 의 최대값을 구하시오.

② 일 때 를 구하시오.

10. 사건 가 독립이고 두 사건 만 발생하는 확률 또는 만 발생하는 확률이 모두 1/4이면 와 를 구하시오.

11. 견습생이 한 기계로 같은 종류의 부속품을 3개 독립적으로 만들 경우 번째 부속품이 불합격품일 확률이 이다. 가 3개 부속품가운데서 합격품의 개수라면 을 구하시오.

12. 모든 고사포들이 비행기를 명중시킬 확률은 0.3이다. 서로 독립적으로 동시에 사격할 때 99%의 확률로 비행기를 명중시키기 위해서는 몇 문의 고사포가 필요한가?

13. 주사위 한개를 던질 때 몇번 던져야 적어도 한번 6점이 나올 확률이 1/2보다 크게 할수 있는가?

14. 한 사격수가 동일한 목표물을 향해 독립적으로 4번 사격한다. 만일 적어도 한발을 명중시킬 확률이 80/81이라면 그 사격수가 한발 쏠 때의 명중률을 구하시오.

15. 매번 사격의 명중률이 0.5일 때 몇번을 사격해야 적어도 한번을 명중시킬 확률이 0.95보다 작아지지 않겠는가?

16. 사냥군이 사냥물과 100 메터 떨어진 곳에서 사냥물을 향해 첫 발을 쏘았다고 가정하면 사냥물을 명중시킬 확률은 0.5이다. 만일 첫 총알이 명중하지 못하면 사냥군은 계속 총을 쏘는데 이때 사냥물과 사냥군의 거리는 이미 150 메터이다. 또 두번째 총알까지 명중하지 못하면 사냥군이 세번째까지 쏘는데 이때 사냥물과 사냥군은 200 메터 떨어졌다. 세번째 총알까지 명중하지 못하면 사냥물이 도망친다. 만일 사냥군이 사냥물을 명중시키는 확률과 거리가 반비례한다면 사냥물이 명중될 확률을 구하시오.

17. 어떤 병원에서 AB형의 혈액이 급히 채혈을 해야 한다. 경험에 의하면 매 100명의 채혈자중 AB형의 혈액은 2명밖에 안된다.

① 20명의 채혈자들가운데서 적어도 1명의 혈액이 AB형일 확률을 구하시오.

② 95%의 확률로 적어도 한명의 AB형혈액을 얻으려면 채혈자가 몇명 있어야 하는가?

18. 사람의 혈액형이 A, B, AB, O형인 확률은 각각 0.37, 0.21, 0.08, 0.34이다. 이제 임의로 4명을 선택하였을 때

① 이 4명의 혈액형이 모두 같지 않은 확률을 구하시오.

② 이 4명의 혈액형이 모두 같을 확률을 구하시오.

19. 두 선수 A, B가 탁구단식경기를 진행하는데 매 경기에서 선수 A, B가 이길 확률은 각각 0.6, 0.4라는것을 알고있다. 3회 2승형식과 5회 3승형식가운데서 어느 경기형식이 A에게 더 유리한가?

20. 세 사람 A, B, C가 경기를 하는데 매 판마다 두 사람씩 경기를하고 여기서 이긴 사람이 세번째 사람과 경기를 진행하는데 이렇게 차례로 경기를 하여 한 사람이 두번 승리할 때까지 계속하며 그 사람이 우승자로 된다. 매 경기에서 쌍방이 이길 확률은 모두 1/2이다. 이제 두 사람 A, B가 먼저 경기를 한다면 매 사람이 우승할 확률은 얼마인가?

21. 두 사람 A, B가 매 회전의 경기에서 이길 확률은 모두 1/2이고 이기면 1점, 지면 0점을 가진다. 누가 먼저 일정한 점수를 따면 승리자로 보고 상품전부를 가지기로 약속하였다. 그러나 경기가 도중에 중단되였을 경우 다음과 같은 상황에서 어떻게 상품을 분배해야 하는가?

① 두 사람 A, B가 각각 회전을 이겨야 승리할수 있다.

② A는 2회전을 더 이겨야 하고 B는 3회전을 더 이겨야 승리할수 있다.

③ A는 화전을 더 이겨야 하고 B는 회전을 더 이겨야 승리할수 있다.

22. 화물차 한대가 물건을 운반하려 간다. 수리공은 운전수에게 6개의 차바퀴가 모두 낡았는데 앞의 2개 바퀴가 파손될 확률은 0.1이고 뒤의 4개 바퀴가 파손될 확률은 0.2라고 알려주었다. 도중에 이 차가 바퀴에 문제가 생겨 고장날 확률이 얼마인가?

23. 일 때 사건 와 가 독립이기 위한 필요충분조건은 이라는것을 증명하시오.

24. , , 이면 사건 와 가 독립이라는것을 증명하시오.

25. 이고 와 가 서로 독립이면 와 가 서로 배반이 아니라는것을 증명하시오.

# 제2장 우연량과 그 분포

정량화된 수학적처리를 진행하기 위해서는 반드시 우연현상의 결과를 수량화해야 한다. 그리하여 우연량을 도입하여 우연현상을 보다 단순하고 통일적으로, 직접적으로 처리할수 있게 한다. 이 장에서는 주로 1차원 우연량과 그 분포에 대하여 론의한다.

## 2.1 우연량과 그 분포

1장에서 우연량을 언급할 때 "우연현상의 결과를 표시하기 위해 리용하는 변수"를 우연량이라고 하였는데 여기서"표시"라는 단어는 무엇을 의미하는가? 이것을 좀 더 깊이 론의한다.

### 2.1.1 우연량의 개념

우연현상에서 많은 표본점은 수량으로 표시되는데 표본점출현의 우연성으로 인하여 그 수량은 우연량으로 나타난다. 실례로

• 주사위를 던질 때 출현하는 점수 는 하나의 우연량이다.

• 매일 어떤 백화점에 입장하는 손님수 , 손님들이 구매한 상품건수 , 손님들이 줄을 서서 기다리는 시간 . 여기서 , , 는 3개의 서로 다른 우연량이다.

• TV의 수명 는 하나의 우연량이다.

• 측정오차 은 하나의 우연량이다.

우연현상들가운데서 적지 않은 표본점은 그 자체로는 수량이 아닌데 이때 연구목적에 따라 우연량을 설정할수 있다. 실례로

• 한 제품을 검사하고 그 제품의 합격여부만 검사하면 그 표본공간은 ={합격품, 불합격품}가 된다. 이때 우연량 를 다음과 같이 설정할수 있다.

|  |  |  |  |
| --- | --- | --- | --- |
| 표본점 | 의 값 | | |
| 합격품 | → | 0 |
| 불합격품 | → | 1 |

여기서 는 "검사한 한개의 제품가운데서 불합격품수"이며 0과 1의 값만 가질수 있다.

만일 이 제품의 불합격률이 라면, 가 취할수 있는 값과 그 확률은 아래와 같다.

|  |  |  |
| --- | --- | --- |
|  | 0 | 1 |
|  |  |  |

• 3개 제품을 검사할 때 8개의 표본점이 있다. 를 "3개 제품중 불합격품수"로 하면 의 값과 표본점사이에는 다음과 같은 대응관계가 있다.

|  |  |  |  |
| --- | --- | --- | --- |
| 표본점 | 의 값 | | |
|  | → | 0 |
|  | → | 1 |
|  | → | 1 |
|  | → | 1 |
|  | → | 2 |
|  | → | 2 |
|  | → | 2 |
|  | → | 3 |

이렇게 가 여러 값들을 취하면 다음과 같은 서로 배반인 사건이 생긴다.



만일 이 제품의 불합격률이 이면 가 매 값을 취할 확률은 다음과 같다.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 0 | 1 | 2 | 3 |
|  |  |  |  |  | |

우연량에 대한 일반적인 정의는 다음과 같다.

정의 2.1.1 표본공간 에서 정의된 실값함수 를 **우연량**이라고 한다. 보통  등 대문자로 우연량을 표시하며 그 값은  등 소문자로 표시한다.

만일 한 우연량이 유한개 또는 셀수 있는 개수의 값만을 가질수 있다면 이것을 **리산우연량**이라고 한다. 또한 어떤 우연량의 가능한 값이 수축의 한 구간 를 채울 경우 이것을 **련속우연량**이라고 한다. 여기서 는 일수 있으며 는 일수 있다.

이 정의에 따르면 우연량 는 표본점 의 함수이며 이 함수는 다른 표본점에 다른 실수를 대응시킬수도 있고 여러 표본점이 같은 실수에 대응될수도 있다. 이 함수의 독립변수(표본점)는 수일수도 있고 수가 아닐 수도 있지만 종속변수는 실수여야 한다.

미적분에서의 변수와 달리 확률론에서 우연량 는 "어떤 분포에 따라 임의의 값을 취하는 변수"이다. 리산우연량을 례로 들면 가 어떤 값을 취할수 있는가뿐아니라 이런 값을 취할 확률이 각각 얼마인가를 알아야 하는데 이것을 위해서는 분포의 개념이 필요하다. 분포의 유무는 일반변수와 우연량을 구분하는 주요한 표징이다.

### 2.1.2 우연량의 분포함수

우연량 는 표본점 의 실값함수이다. 만일 가 실수부분모임(즉 )이면 은 우연사건 을 나타낸다. 특히 사건을 표시하기 위해 우연량 와 일부 실수를 같기부호나 안같기부호로 련결한다. 실례로 는 모두 우연사건이다. 구체적으로

• 를 주사위를 한번 던질 때 출현하는 점수라고 하면 가 취할수 있는 가능한 값은 1, 2, …, 6이다. 이것은 리산우연량이다. 사건 ="점수가 3보다 작거나 같다"는 으로 표시될수 있다.

• 를 하루에 어떤 상점에 오는 손님수라고 하면 의 가능한 값은 0, 1, 2,…, , …이므로 역시 리산우연량이다. 사건 ="최소 1000명의 손님이 온다"는 으로 표시된다.

• 를 전기제품의 사용수명이라고 하면 의 가능한 값은 구간 이다. 이것은 련속우연량이다. 사건 ="사용수명이 40000~50000시간이다"는 로 표시될수 있다.

의 통계적법칙성을 파악하기 위해서는 가 매 값을 취할 확률을 알기만 하면 된다.

이므로 임의의 실수 에 대하여 사건 의 확률을 알면 되는데 그 확률은 루적의 특성을 가지고있다. 일반적으로 로 표시한다. 그리고 이 확률 는 와 관계가 있고 마다 이 루적확률의 값이 다르다. 따라서 라고 표시한다. 가 모든 에 대하여 다 정의되므로 는 정의역이 이고 값구역이 [0, 1]인 함수이다. 이것이 바로 분포함수이다.

정의 2.1.2 가 우연량일 때 임의의 실수 에 대하여



를 우연량 의 **분포함수**라고 한다. 또는 가 에 따른다고 말하며 로 표시한다. 때로는 로 의 분포함수라는것을 나타내기도 한다.

실례 2.1.1 반경이 인 원안에 임의로 한 점을 던져 이 점에서 중심까지의 거리 의 분포함수 를 구하고 를 구하시오.

풀기. 사건 ""는 던진 점이 반경 인 원안에 있다는것을 표시하므로 기하확률로부터 이고 일 때 이고 이면 이다.

따라서 .

분포함수의 정의에서 알수 있듯이 우연량 (리산형 또는 련속형)는 모두 하나의 분포함수를 가진다. 이 분포함수를 리용하여 우연량 와 관련된 사건의 확률을 계산할수 있다. 아래에서 먼저 분포함수의 세가지 기본성질을 증명하다.

정리 2.1.1 모든 분포함수 는 다음과 같은 세가지 기본성질을 가진다.

① 단조성: 는 전체 실수축 우에서 정의된 단조비감소함수이다. 즉 임의의 에 대하여 이다.

② 유계성: 임의의 에 대하여 이고 또한



③ 오른쪽련속성: 는 의 오른쪽련속함수로서 임의의 에 대하여 . 즉 .

증명. ①은 분명하므로 ②를 증명하자. 는 사건 의 확률이므로 이다. 의 단조성으로부터 임의의 에 대하여 이 모두 존재한다. 또한 확률의 셀수 있는 가법성으로부터



이로부터 임을 알수 있다.

다시 ③을 증명하자. 는 단조유계비감소함수이므로 임의의 한 점 의 오른쪽극한 은 반드시 존재한다. 오른쪽련속성을 증명하려면 임의의 단조감소수렬 에 대하여 일 때 가 성립한다는것을 증명하면 된다.



이므로 이다.

이로써 세가지 기본성질이 전부 증명되였다.

이 세가지 기본성질은 분포함수가 반드시 가져야 할 성질이다. 또한 이 세가지 성질을 만족하는 함수는 반드시 어떤 우연량의 분포함수임을 증명할수 있다. 따라서 이 세가지 기본성질은 어떤 함수가 분포함수로 될수 있는가, 없는가를 판별하는 필요충분조건으로 된다.

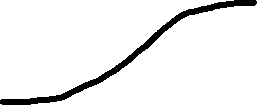
우연량 의 분포함수가 있으면 와 관계되는 여러가지 사건들의 확률을 모두 분포함수로 편리하게 표시할수 있다. 례컨대 임의의 실수 와 에 대하여



특히 가 와 에서 련속일 때 이다.

이러한 공식은 앞으로의 확률계산에서 자주 보게 된다.

실례2.1.2 함수 는 전체 수축에서 련속이고 엄격히 단조증가하며 이다. 이 가 분포함수의 세가지 기본성질을 모두 만족하므로 는 하나의 분포함수이며 이 분포함수를 **코씨분포함수**라고 한다. 그 그라프를 그림 2.1.1에 보여주었다.













1

-1

그림 2.1.1 코씨분포함수

만일 가 코씨분포에 따른다면

.

### 2.1.3 리산우연량의 확률분포렬

리산우연량의 분포는 일반적으로 아래와 같이 정의되는 분포렬로 표현된다.

정의 2.1.3 리산우연량 의 가능한 값들이 모두 라고 하면 가 를 취할 확률



을 의 **확률뷴포렬** 또는 간단히 **분포렬**이라고 하고 로 표시한다.

분포렬은 다음의 표형식으로도 나타낼수 있다.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  | … |  | … |
|  |  |  | … |  | … |

또는 로 표시할수도 있다.

제1장에서 우리는 이미 여러 개의 분포렬을 보았다. 매개 리산우연량은 서로 다른 분포렬을 가질수 있으며 심지어 한 표본공간우에서도 몇가지 서로 다른 분포렬에 따르는 우연량을 정의할수 있다. 구체적인 실례를 들어보자.

실례 2.1.3 두개의 주사위를 던지는데 그 표본공간 는 36개의 등가능한 표본점을 포함한다. . 우에서 다음과 같은 3개의 우연량 를 정의하자.

• 는 점수들의 합으로서 가능한 값은 2, 3, …, 12 등 모두 11개의 값이다. 그 정의는 그림 2.1.2 (a)를 보시오.

• 는 6점인 주사위의 개수로서 가능한 값은 0, 1, 2 등 3개의 값을 가질수 있다. 이에 대한 정의는 그림 2.1.2 (b)를 보시오.

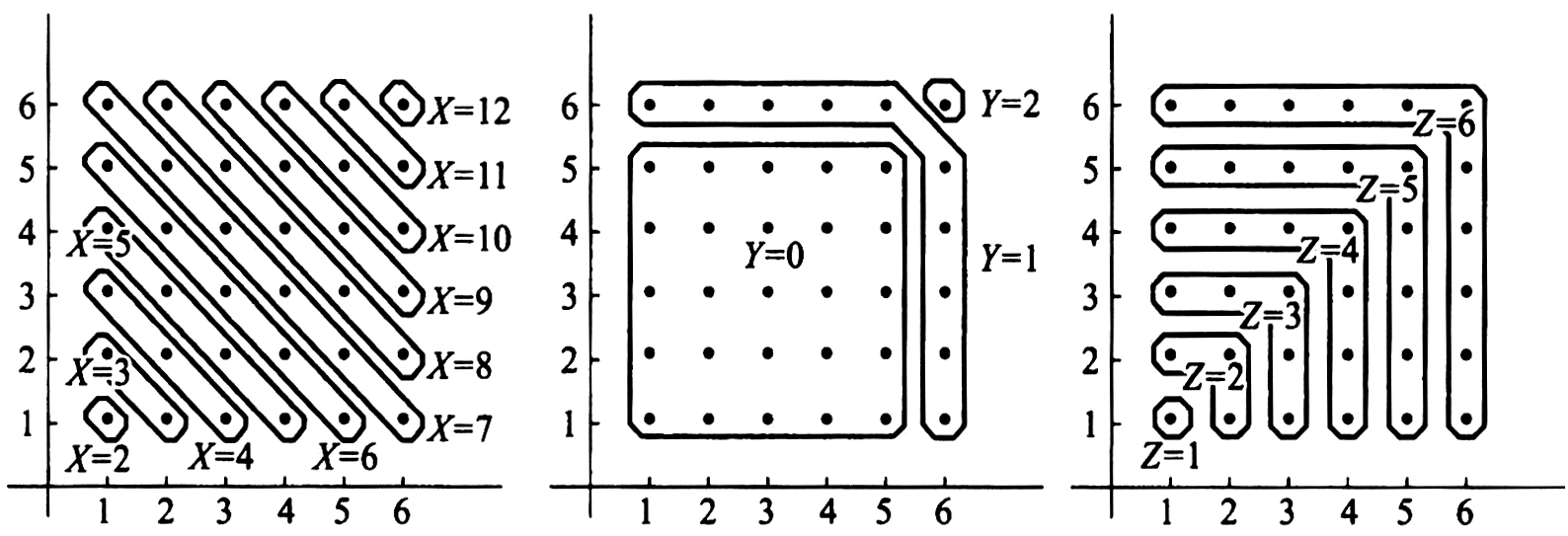
• 는 최대점수로서 가능한 값은 1, 2, …, 6과 같은 6개이다. 그 정의는 그림 2.1.2 (c)를 보시오.

(a) 점수의 합 의 정의

(b) 6점의 개수 의 정의

(c) 최대점수 의 정의

그림 2.1.2 동일한 표본공간우에서의 서로 다른 우연량



이 3개 우연량의 분포렬을 고전적인 방법으로 계산하면 다음과 같다.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|  |  |  |  |  |  |  |  |  |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
|  | 0 | 1 | 2 |
|  |  |  |  |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | 1 | 2 | 3 | 4 | 5 | 6 |
|  |  |  |  |  |  |  |

마찬가지로 이 표본공간우에서 다른 리산우연량들도 정의될수 있다.

성질 2.1.1분포렬의 기본성질

① **비부성 **.

② **정규성 **.

우의 두가지 기본성질은 분포렬이 반드시 만족해야 하는 성질이며 또 어떤 수렬이 분포렬이 될수 있는가를 판단하는 필요충분조건이다.

리산우연량 의 분포함수는 의 분포렬을 리용하여 표시할수 있다. . 그 그라프는 계단함수의 유한합 (또는셀수 있는 무한합)인데 구체적으로는 아래의 실례를 참고할수 있다. 그러나 리산의 경우 일반적으로 분포렬을 리용하여 그 분포를 서술하며 분포함수는 거의 리용되지 않는다. 그것은 리산우연량 와 관련된 사건의 확률을 구할 때 분포함수보다 분포렬을 리용하는것이 더 편리하기때문이다.

실례2.1.4 리산우연량 의 분포렬이 다음과 같다고 하자.

|  |  |  |  |
| --- | --- | --- | --- |
|  | -1 | 2 | 3 |
|  | 0.25 | 0.5 | 0.25 |

이때 를 구하고 의 분포함수를 써보시오.

풀기.



그림 2.1.3에 보여준것처럼 의 그라프는 계단모양의 곡선인데 의 가능한 값 -1, 2, 3에서 오른쪽련속인 도약점이 있으며 그 도약의 크기는 각각  가 가능한 값에서의 확률이다.

특히 상수 는 하나의 값만을 취하는 우연량 로 볼수 있다. .

이런 분포를 흔히 단일점분포 또는 퇴화분포라고 하는데 그 분포함수는 다음과 같다.



그 그라프는 그림 2.1.4에 보여주었다.

그림 2.1.3 리산우연량의 분포함수

그림 2.1.4 단일점분포함수



1

-1



1

2

3









1



아래의 례는 구체적으로 리산우연량 의 분포렬을 구할 때 중요한것은 의 모든 가능한 값과 이러한 값을 취할 확률을 구하는것이라는것을 설명하여준다.

실례 2.1.5 자동차가 한 도로를 따라 달릴 때 빨간 신호등과 푸른 신호등이 설치되여있는 3개의 교차점을 지나야 한다. 매개 신호등이 빨간 신호와 푸른 신호를 나타내는 시간이 같고 서로 독립적으로 동작한다고 가정할 때 이 자동차가 처음 빨간 신호등을 만날 때까지 이미 지나간 교차점의 수 의 확률분포를 구하시오.

풀기. 의 가능한 값은 0, 1, 2, 3이다. ="자동차가 째 교차점에서 빨간 신호등을 만난다"라고 하면 은 서로 독립이고 이므로



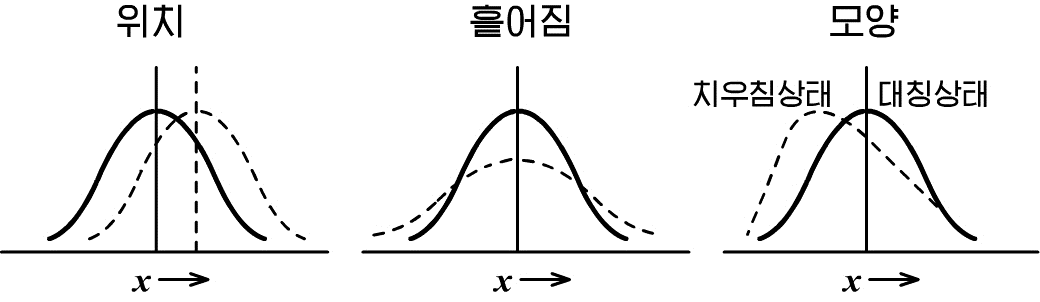
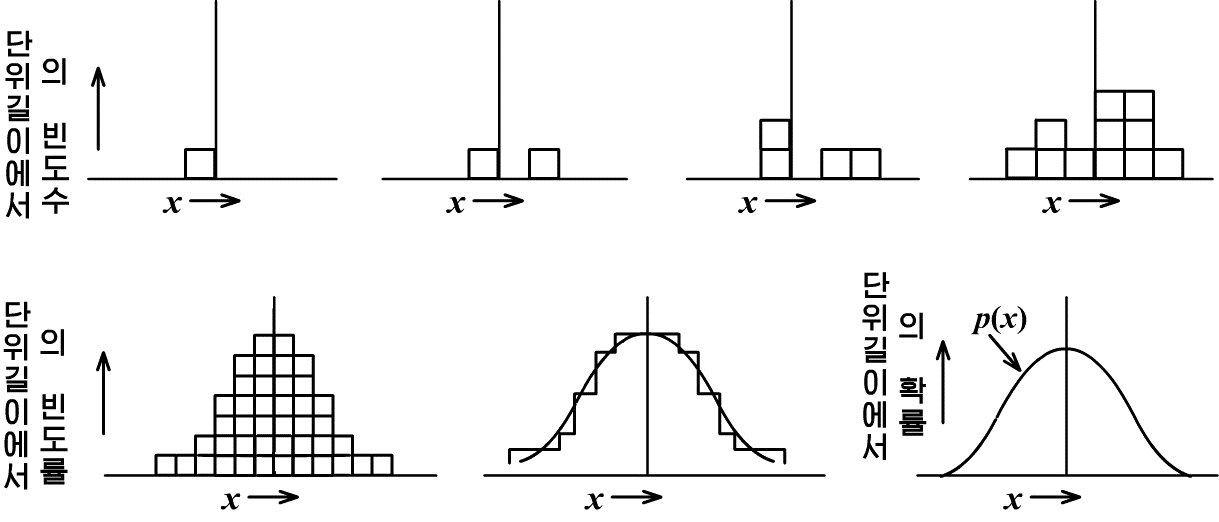
따라서 의 분포렬은 다음과 같다.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 0 | 1 | 2 | 3 |
|  | 0.5 | 0.25 | 0.125 | 0.125 |

### 2.1.4 련속우연량의 확률밀도함수

앞에서 우리는 련속우연량을 직관적으로 묘사하였다. 즉 가능한 모든 값은 어떤 구간 를 가득 채우지만 이 구간안에는 무한히 많은 셀수 없는 개수의 실수가 있으므로 이러한 우연량의 확률분포는 분포렬의 형식으로 표현할수 없고 대신 확률밀도함수로 표현해야 한다. 아래에서 한가지 례를 들어 확률밀도함수의 유래를 알아보기로 하자.

실례2.1.6 가공한 기계축의 직경을 측정한 값 는 하나의 우연량이다. 축의 직경을 측정하여 그 측정값 를 수축에 표식하면 많은 가 축적될 때 일정한 도형이 형성된다. 안정한 그라프를 얻기 위하여 세로축을 "단위길이에서의 빈도수"가 아니라 "단위길이에서의 잦음률"로 설정하였다. 잦음률의 안정성때문에 이 그라프는 측정값 의 수가 많을수록, 단위길이가 작을수록 더 안정하며 미끄러운 곡선으로 나타난다(그림 2.1.5 (a)). 이때 이 곡선의 세로좌표는 "단위길이에서의 확률"이며 단위길이가 0에 가까워지면 세로좌표는 "한 점에서의 확률밀도"가 된다. 그러면 이 곡선이 표시하는 함수 를 확률밀도함수라고 하며 이것은 가 "어떤 곳(실례로 가운데) 에서 값을 가질 기회가 많고 다른 곳(실례로 량끝)에서는 기회가 작다"라는 통계적규칙성을 나타낸다. 확률밀도함수 는 여러 형태로 표시되는데 위치가 다르거나 흩어짐정도가 다르고 또는 모양이 다를수 있다(그림 2.1.5(b)).



(a) 확률밀도함수 의 형성과정

(b) 확률밀도함수 의 여러가지 모양

그림 2.1.5

이것은 서로 다른 우연량이 취하는 값이 통계적법칙성에서 차이난다는것을 반영하고있다.

확률밀도함수 의 값은 확률이 아니지만 단위미분 를 곱하면 소구간 에서의 확률에 대한 근사값을 얻을수 있다. 즉 이다.

에서 린접한 많은 단위미분들의 루적을 통하여 의 에서의 적분을 얻을수 있는데 이 적분값이 바로 가 에서 값을 취할 확률이다. 즉 .

특히 에서의 의 적분은 바로 분포함수 이다. .

이 관계는 련속우연량 에 대한 확률밀도함수 의 가장 본질적인 속성이다. 이 모든 연산은 가 비부인 적분가능한 함수일것을 요구한다. 그러므로 확률밀도함수 는 다음과 같이 엄밀하게 정의된다.

정의 2.1.4 우연량 의 분포함수가 라고 가정한다. 만일 실수축우에서 부가 아닌 적분가능한 함수 가 존재하여 임의의 실수 에 대하여



을 만족한다면 를 의 **확률밀도함수**라고 하며 간단히 **밀도함수** 또는 **밀도**라고 한다. 동시에 를 **련속우연량**이라고 하고 를 **련속분포함수**라고 한다.

이렇게 련속우연량과 그 분포에 대한 정의를 완성하였다. 식 (2.1.4)로부터 의 도함수가 존재하는 점에서



라는것을 알수 있다. 는 (루적)확률함수이고 그 도함수 는 확률밀도함수이며 이로부터 가 확률밀도함수로 불리우는 리유를 알수 있다.

식 (2.1.5)을 리용하여 분포함수로부터 밀도함수를 구할수 있다. 례를 들면 실례 2.1.2에 제시된 코씨분포함수는 모든 점에서 미분가능하므로 코씨분포의 밀도함수는 이다.

성질 2.1.2 밀도함수의 기본성질

① 비부성 .

② 정규성 . (의 적분가능성을 포함한다)

이 두가지 기본성질은 밀도함수가 반드시 만족하여야 할 성질로서 어떤 함수가 밀도함수가 될수 있는가를 판별하는 필요충분조건이다. 실례로 어떤 함수 가 밀도함수라는것이 알려지고 에 어떤 미지의 상수가 있다면 정규성 을 리용하여 그 상수의 값을 구할수 있다.

실례2.1.7 구간 에서 임의로 점을 선택하고 그 점의 좌표를 로 표시한다. 이 점이 의 임의의 부분구간에 놓일 확률은 이 소구간의 길이에 정비례하며 소구간의 위치와는 관계없다고 가정한다. 의 분포함수와 밀도함수를 구하시오.

풀기. 의 분포함수를 라고 하면

일 때 는 불가능한 사건이므로 이다.

일 때 는 필연사건이므로 이다.

일 때 인데 여기서 는 비례곁수이다. 이므로 이다.

따라서 의 분포함수는 이다.

이제 의 밀도함수 를 구한다.

 또는 일 때 이고 일 때 이다. 또한 에서 는 임의의 값을 취할수 있으며 일반적으로 가까운 곳에서 값을 취하면 좋다. 이것은 확률 (적분)의 계산에 영향을 주지 않는다. 따라서 의 밀도함수는 다음과 같다.



이 분포는 구간 에서의 평등분포로서 로 표시하며 그 밀도함수 와 분포함수 는 그림 2.1.6에 보여주었다.

1. 의 그라프
2. 의 그라프

그림 2.1.6 우에서의 평등분포



















1

사실 이 실례는 바로 1장에서 언급한 기하확률인데 이렇게 기하확률과 평등분포사이의 련계를 확립할수 있다.

아래에서 련속우연량의 실례를 들어보자.

실례 2.1.8 어떤 종류의 전자부속품의 수명 (시간단위)는 다음과 같은 확률밀도함수를 가진다.



이제 서로 독립적으로 동작하는 이런 부속품들이 많을 때 다음의 물음에 대답하시오.

① 임의로 하나를 선택할 경우 그 수명이 1,500시간이상일 확률은 얼마인가?

② 임의의 4개를 선택할 경우 4개의 수명이 모두 1,500시간이상일 확률은 얼마인가?

③ 임의의 4개를 선택할 경우 4개중 적어도 1개의 수명이 1,500시간이상일 확률은 얼마인가?

④ 만일 어떤 부속품의 수명이 1,500시간이상이라는것을 알고 있는 경우 그 부속품의 수명이 2,000시간이상인 확률은 얼마인가?

풀기. 먼저 의 분포함수를 계산한다. .

① .

② 매 소자의 작업이 독립적이기때문에 구하려는 확률은 이다.

③ 구하려는 확률은 다음과 같다.

(4개중 적어도 1개의 수명이 1500시간이상) = 1-(4개 부속품의 수명이 모두 1500시간보다 작거나 같다).

④ 이것은 조건부확률 를 구하는것이다. 라고 하면 이고 이므로 .

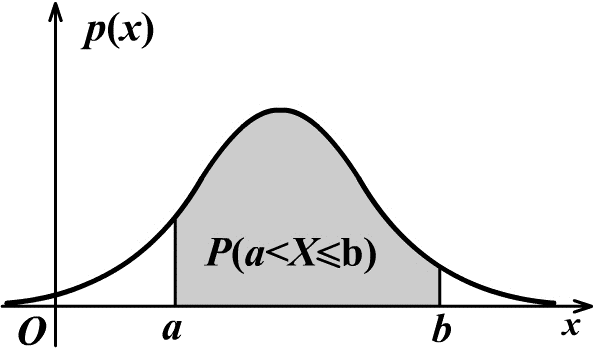
아래에서 밀도함수와 분포렬의 차이점과 공통점에 대해 설명하자.

리산우연량의 경우 , 여기서 는 가 취할수 있는 값이다.

그리고 련속우연량의 경우 이다. 그 의미는 그림 2.1.7에서 구간 우의 곡변제형면적이다.

이런 의미에서 확률밀도함수와 확률분포렬의 작용은 비슷하지만 그 차이도 명백하다.

그림 2.1.7 련속우연량인 경우의 확률



① 리산우연량의 분포함수 는 항상 오른쪽련속인 계단함수이고 련속우연량의 분포함수 는 반드시 전체 수축에서 련속인 함수이다. 전자는 이미 앞에서 설명하였고 후자는 임의의 점 에서의 증분 에 대응한 분포함수의 증분에 대하여 이기때문이다.

② 리산우연량 는 가능한 값 에서의 확률이 0이 아니지만 련속우연량 는 우의 임의의 점 에서의 확률이 항상 0이다. 그것은 이기때문이다.

이것은 불가능한 사건의 확률이 0 이지만 확률이 0인 사건 (례하면 )은 반드시 불가능한 사건이 아니라는것을 보여준다. 마찬가지로 필연사건의 확률은 1이지만 확률이 1인 사건이 반드시 필연사건인것은 아니다.

③ 련속우연량  가 한 점만을 취할 확률은 항상 0이므로 사건 에서 나 를 빼버리는것은 그 확률에 영향을 주지 않는다. 즉 .

이것은 계산에 큰 편리를 가져다준다. 그러나 리산우연량의 경우 이 성질은 성립하지 않으며 매 점에 대하여 계산해야 한다.

④ 일부 점들에서 밀도함수 의 값이 변해도 적분값에는 영향을 주지 않으며 따라서 그 분포함수 의 값에도 영향을 주지 않는다. 이것은 어떤 련속분포의 밀도함수가 유일하지 않다는것을 의미한다. 례를 들면 실례2.1.7에서 과 에서 의 값을 다음과 같이 변경하여보자.



두 함수 는 모두 우에서의 평등분포에 대한 밀도함수이지만 자세히 고찰해보면 이다.

이로부터 두 함수가 확률의 의미에서는 차이가 없다는것을 알수 있다. 이때 는 "거의 모든 곳에서 같다"라고 한다.

그 의미는 확률론에서 확률이 0인 사건을 제외한 다음에도 두 함수의 동등성과 기타 임의의 문제를 론할수 있다는것이다. 이것이 바로 확률론이 미적분과 다른 점이고 확률론의 매력이기도하다.

리산분포와 련속분포외에도 다른 형태의 분포가 있는데 아래의 실례를 보면 알수 있다.

실례 2.1.9 다음과 같은 함수 는 그라프가 그림 2.1.8과 같은 하나의 분포이다.



그림 2.1.8 리산분포도 아니고 련속분포도 아닌 분포함수의 실례



1

0.5

1





그림에서 볼수 있다싶이 이 함수는 계단함수도 아니고 련속함수도 아니다. 그러므로 리산분포도 아니고 련속분포도 아닌 새로운 종류의 분포이다. 이런 분포함수 는 일반적으로 두개 분포함수의 불룩결합으로 분해한다. 실례로 실례 2.1.9의 분포함수는 다음과 같이 분해될수 있다.

.

여기서 

는 (리산)단일점분포함수이고 는 (련속)평등분포 의 분포함수이다.

련습문제 2.1

1. 주머니속에 번호가 1, 2, 3, 4, 5인 공 5개가 있을 때 그중에서 임의로 3개를 취한 공들가운데서 가장 큰 번호를 로 표시한다.

① 의 분포렬을 구하시오.

② 의 분포함수를 구하고 그라프를 그리시오.

2. 한개의 주사위를 두번 던질 때 다음과 같은 우연량의 분포렬을 구하시오.

① 는 두번의 던지기에서 얻은 최소점수를 표시한다.

② 는 두번의 던지기에서 얻은 점수차의 절대값을 표시한다.

3. 주머니에 흰 공 7개, 검은 공 3개가 들어 있다.

① 매번 임의로 공을 하나 꺼내고 다시 넣지 않는다. 이때 흰공을 처음 꺼낼 때까지 공을 꺼낸 차수 의 확률분포렬을 구하시오.

② 만일 꺼낸 공이 검은 공이면 다시 넣지 않고 흰 공을 한개 넣을 때 의 확률분포렬을 구하시오.

4. 세개의 상자가 있는데 첫 번째 상자에는 흰 공 1개와 검은 공 4개가, 두번째 상자에는 흰 공 2개와 검은 공 3개가, 세번째 상자에는 흰 공 3개와 검은 공 2개가 들어 있다. 이제 임의로 상자를 하나 선택하고 그 안에서 공 3개를 임의로 꺼낸다. 로 꺼낸 흰 공의 수를 표시한다.

① 의 확률분포렬을 구하시오.

② 꺼낸 흰 공의 수가 2보다 적지 않을 확률은 얼마인가?

5. 한 주사위를 4번 던져 6점이 나타나는 차수의 확률분포를 구하시오.

6. 52장의 주패 한조에서 임의로 5장을 뽑을 때 그가운데서 스페이드의 개수에 대한 확률분포를 구하시오.

7. 제품더미에 100개의 제품이 있는데 그중 10개가 불합격품이다. 검수규칙에 따라 그중 5개 제품을 취하여 품질검사를 한다. 만일 5개가운데 불합격품이 없으면 이 제품더미는 합격되고 그렇지 않으면 이 제품들을 다시 하나하나 검사해야 한다.

① 5개가운데 불합격품수 의 분포렬을 구하시오.

② 이 제품들을 하나하나 검사해야 할 확률은 얼마인가?

8. 우연량 의 분포함수를 다음과 같이 가정한다.



이때 의 확률분포렬과 을 구해보시오.

9. 우연량 의 분포함수를 다음과 같이 가정한다.



이때 를 구해보시오.

10. 인데 그중 이다. 를 구하시오.

11. 5개의 수자 1, 2, 3, 4, 5에서 임의로 세개를 취하고 크기순서로 배렬한 다음 라고 표시한다. 이라고 할 때

① 의 분포함수을 구하시오.

② 과 을 구하시오.

12. 우연량 의 밀도함수를 다음과 같이 가정한다.



이때 의 분포함수를 구하시오.

13. 우연량 의 밀도함수는 다음과 같다.



이때 를 구하시오.

14. 우연량 의 밀도함수를 다음과 같이 가정한다.



① 곁수 를 구하시오.

② 가 구간 에 놓일 확률을 구하시오.

15. 련속우연량 의 분포함수를 다음과 같이 가정한다.



① 곁수 를 구하시오.

② 가 구간 (0.3, 0.7)에 놓일 확률을 구하시오.

③ 의 밀도함수를 구하시오.

16. 학생이 숙제를 완성하는 시간 는 우연량이며 그 단위는 시간으로서 밀도함수는 다음과 같다.



① 상수 를 결정하시오.

② 의 분포함수를 쓰시오.

③ 20분내에 숙제를 끝낼 확률을 구하시오.

④ 숙제를 끝내는데 10분이상 걸릴 확률을 구하시오.

17. 어떤 연유공급소에서는 매주 한번씩 연유를 보충한다. 공급소의 주간 판매량(단위:키로리터)이 우연량이라고 하고 그 밀도함수는 다음과 같다고 하자.



일주일내에 연유가 떨어질 확률을 5%이하로 하려면 이 공급소의 탕크는 얼마나 커야 하는가.

18. 우연량 와 의 분포가 같다고 가정하고 의 밀도함수는 다음과 같다.



사건 과 가 독립이고 라는것을 알고 상수 를 구하시오.

19. 련속우연량 의 밀도함수 가 짝함수이고 가 의 분포함수이면 임의의 실수 에 대하여 다음과 같은 관계가 성립한다는것을 증명하시오.

① 

② 

③ 

## 2.2 우연량의 수학적기대값

우리가 알고 있다싶이 모든 우연량에는 하나의 분포 (분포렬, 밀도함수, 분포함수)가 있으며 우연량마다 분포가 다를수도 있고 같을수도 있다. 분포는 우연량이 취하는 값의 통계적규칙성을 전반적으로 묘사하며 분포로부터 관련있는 임의의 사건의 확률을 얻을수 있다. 이외에도 분포로부터 대응한 우연량의 평균, 분산, 분위수 등과 같은 특징수들을 계산할수 있다. 이러한 특징수들은 각각 일정한 측면에서 분포의 특징을 묘사한다. 례를 들어 갓난아이의 체중은 우연량이며 그 평균값은 체중의 한 측면을 나타낸다. 우연량의 분포를 알고있을 때 그 평균을 어떻게 구할것인가 하는것을 이 절에서 론의한다.

이 절에서는 우연량의 가장 중요한 특징수인 수학적기대값에 대해 취급한다.

### 2.2.1 수학적기대값의 개념

일상생활에서 "기대"는 마음속으로 바라는 소원을 가리키는데 확률론에서 수학적기대값은 력사상 유명한 도박밑천의 분할문제에서 기원되였다.

실례 2.2.1(도박밑천의 분할문제) 17세기 중엽 한 도박군이 프랑스의 수학자 파스칼(1623-1662)에게 오래동안 생각해 온 도박밑천의 분할문제를 제기했다. 즉 두명의 도박군 A, B가 서로 실력이 비슷한데 각각 50프랑을 걸고 게임을 하며 매판에 꼭 승부가 난다. 그들은 세 판을 먼저 이기는 사람에게 100프랑을 모두 주기로 했다. A가 두 판을 이기고 B가 한 판을 이겼을 때 어떤 사정이 있어 도박을 중지해야 한다. 이때 100프랑을 어떻게 나누어야 공평하다고 할수 있는가?

이 문제는 적지 않은 사람들의 흥미를 불러일으켰다. 우선 사람들은 똑같이 나누는것은 A에게 불공평하고 전부 A에게 주는것은 B에게 불공평하다는것을 인식하였다. 일정한 비률에 따라 A에게 좀 많이 분배하고 B에게 좀 적게 분배하는것이 합리한 방법이다. 이때 어떤 비률로 나누겠는가 하는 문제가 제기된다. 다음과 같은 두가지 분할방법이 있다.

① A는 100프랑중 3분의 2를 가지고 B는 3분의 1을 가진다. 이것은 이긴 판수에 기초한것이다.

② 1654년에 파스칼은 다음과 같은 분할법을 제기하였다. 다시 도박을 계속한다고 하면 A가 최종적으로 가지는 는 하나의 우연량인데 그 값은 0 또는 100일수 있다. 다시 두판을 두면 반드시 결판이 날것이고 그 결과는 네가지 경우 AA, AB, BA, BB중 어느 하나일것이다.

여기서 "AB"은 첫판은 A가 이기고 두번째 판은 B가 이긴다는것을 의미한다. 이 네가지 경우가운데서 세가지는 A가 100프랑을 가지는 경우이고 오직 한가지 경우 "BB"인 경우에만 A가 0프랑을 가진다. 두 선수의 실력이 엇비슷하기때문에 A가 100프랑을 가질 가능성은 3/4 이고 0프랑을 가질 가능성은 1/4이다. 즉 의 분포는 다음과 같다.

|  |  |  |
| --- | --- | --- |
|  | 0 | 100 |
|  | 0.25 | 0.75 |

우의 분석을 통하여 파스칼은 A가 "기대"하는 소득은 프랑이라고 주장하였다. 즉 A는 75 프랑, B는 25프랑을 가진다. 이러한 분류방법은 이미 진행한 경기점수를 고려하였을뿐만아니라 다시 경기를 진행해나가는데 대한 일종의 "기대"도 포함하였기때문에 ①의 분할방법보다 더 합리적이다.

수학적기대값이라는 이름은 이렇게 유래되였는데 사실 이 이름은 "평균값"이라고 하는것이 더 리해하기 쉽다. 즉 다시 도박을 하면 A가 "평균" 75프랑을 가진다.

이제 분포로부터 어떻게 "평균값"을 구하는가를 단계적으로 분석해보기로 하자.

① 산수평균: 개의 수 가 있다고 할 때 이 개 수의 산수평균을 구하는것은 간단하다. 모두 더하여 으로 나누면 된다.

② 무게평균: 개의 수가운데 같은것이 있다면 일반성을 잃지 않고 그중 개가 라고 할수 있다. 이때

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 값 |  |  | … |  |
| 빈도수 |  |  | … |  |
| 잦음률 |  |  | … |  |

그 "평균값"은 다음과 같다. 

사실 이 무게평균의 "무게" 는 바로 가 나타나는 잦음률로서 이 크면 확률로 안정된다.

③ 리산우연량 에 대하여 가능한 값이 이라고 하자. 만일 이 개의 수를 더한 후 으로 나누어 "평균값"으로 삼는다면 이것은 확실히 잘못된것이다. 그 원인은 가 매 값을 취할 확률이 일반적으로 다르고 확률이 큰것이 나타날 기회가 크므로 계산에서 그 무게도 커야 하기때문이다. 우의 실례에서 본 도박밑천의 분할문제는 취할수 있는 값의 확률을 일종의 "무게"로 하여 무게평균을 구하는것이 아주 합리하다는것을 시사하고있다.

이상의 분석을 통하여 수학적기대값의 정의를 내릴수 있다.

### 2.2.2 수학적기대값의 정의

정의 2.2.1 리산우연량 의 분포렬이 이라고 하자.

만일 이면



을 우연량 의 **수학적기대값** 또는 그 분포의 수학적기대값이라고 하며 간단히 **기대값**, **평균값**이라고 부른다. 만일 합렬 이 수렴하지 않으면 의 수학적기대값은 존재하지 않는다.

우의 정의에서 무한합렬의 절대수렴을 요구하는 목적은 유일한 수학적기대값을 가지게 하려는데 있다. 우연량의 값은 정일수도 있고 부일수도 있으며 그 값을 취하는 앞뒤순서도 정해져있지 않으므로 무한합렬의 리론으로부터 알수 있다싶이 만일 이 무한합렬이 절대수렴하면 그 합은 순서의 변화에 대한 영향을 받지 않는다. 유한항의 합은 순서의 변화에 영향을 받지 않으므로 유한개의 가능한 값을 가지는 우연량은 수학적기대값이 항상 존재한다.

련속우연량의 수학적기대값의 정의와 의미는 리산우연량의 경우와 완전히 비슷한데 리산합을 적분으로 바꾸기만 하면 된다.

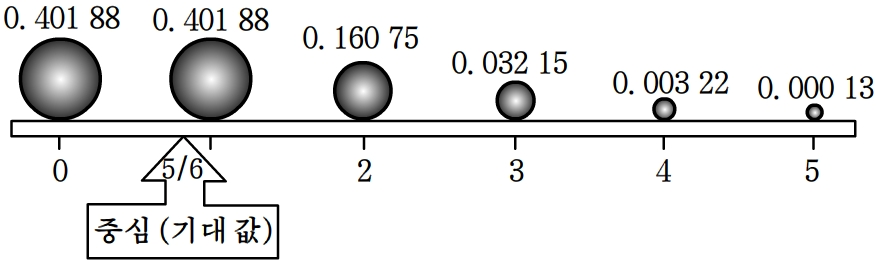
정의 2.2.2 련속우연량 의 밀도함수를 라고 하자. 만일 이면



를 의 수학적기대값 또는 분포 의 수학적기대값, 기대값, 평균값이라고 부른다. 만일 이 수렴하지 않으면 는 수학적기대값을 가지지 않는다.

수학적기대값 의 물리적의미는 무게중심이다. 확률 를 점 우의 질량으로, 확률분포를 축우에서 질량의 분포로 본다면 의 수학적기대값 은 그림 2.2.1에서 볼수 있다싶이 그 질량분포의 중심위치이다.

그림 2.2.1 확률질량모형: 5개의 주사위를 동시에 던졌을 때 6점이 출현하는 개수 의 수학적기대값 이 바로 중심위치



수학적기대값은 리론적으로 중요한 의의를 가지며 그것은 우연성을 제거하는 주요한 수단이다.

수학적기대값을 현실에서 널리 응용하고있는데 는 흔히 의 분포에 대한 대표값(일종의 통계지표)으로서 같은 종류의 지표를 비교하는데 참가한다. 아래의 실례는 수학적기대값이 널리 응용된다는것을 보여준다.

실례 2.2.2 명의 집단에서 어떤 질병을 조사하기 위해서는 그들에 대한 혈액검사를 해야 한다.매 사람의 피를 따로 검사한다면 총 번 검사해야 한다. 작업량을 줄이기 위해 한 통계학자가 다음과 같은 방법을 제기하였다. 명을 한개 조로 하여 조를 가르고 같은 조의 명에 대한 혈액표본을 혼합하여 검사한다. 만일 혼합된 혈액표본이 음성을 보이면 이 명의 혈액이 모두 음성이라는것으로서 질병과 관련이 없으며 결국 명의 검사를 한번만 하면 되므로 매 사람에 대하여 번 검사를 한것에 해당한다. 만일 혼합된 혈액표본이 양성을 보이면 명중에 적어도 한명이 양성이므로 다시 이 명의 혈액을 검사한다. 이때 이 명에 대하여 번 검사하므로 매 사람에 대하여 번 검사하는것에 해당한다. 이 질병의 발병률이 이고 이 질병에 걸리는것이 사람마다 서로 독립이라고 가정할 때 이런 검사방법으로 평균검사회수를 줄일수 있는가?

풀기. 를 이 집단의 매 사람에 대하여 필요한 혈액검사회수라고 하면 의 분포렬은

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  |  |  |

그러므로 매 사람당 혈액검사차수는 다음과 같다.

.

따라서  즉 이 만족하는 를 선택하면 혈액검사회수를 줄일수 있으며 를 적당히 선택하여 를 최소화할수 있다. 례를 들어 일 때 에 따르는 의 값을 표 2.2.1에 주었다. 표에서 볼수 있듯이 일 때 평균혈액검사회수가 1을 초과하는데 이것은 개별적을 검사하는 작업량보다 크다. 일 때 평균혈액검사화수는 각이한 정도로 감소되는데 특히 일 때 평균혈액검사회수가 최소이고 작업량도 40% 감소된다.

표 2.2.1 일 때 의 값

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 2 | 3 | 4 | 5 | 8 | 10 | 30 | 33 | 34 |
|  | 0.690 | 0.604 | 0.594 | 0.610 | 0.695 | 0.751 | 0.991 | 0.9994 | 1.0016 |

또한 서로 다른 발병률 에 대해 한개 그룹의 최량한 인원수 를 계산할수 있다. 표 2.2.2로부터 발병률 가 작을수록 그룹검정의 효과가 크다는것을 알수 있다. 실례로 일 때 11명을 한그룹으로 하여 혈액검사를 하면 혈액검사량이 80%정도 줄어든다. 이것은 제2차 세계대전 당시 신병모집을 대량으로 할 때 혈액검사를 간소화하기 위한 조치였다.

표 2.2.2 서로 다른 발병률 에 대한 최량그룹인원수 과 그 

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | 0.14 | 0.10 | 0.08 | 0.06 | 0.04 | 0.02 | 0.01 |
|  | 3 | 4 | 4 | 5 | 6 | 8 | 11 |
|  | 0.697 | 0.594 | 0.534 | 0.466 | 0.384 | 0.274 | 0.196 |

실례 2.2.3 복권은 장당 판매가격이 5원이며 각각 채권번호가 1개씩 있다. 100만장을 판매할 때마다 추첨을 진행하는데 관중앞에서 6자리의 당첨번호 (000000부터 999999까지의 모든 수가 출현할수 있다)를 추첨기로 뽑는다. 추첨규칙은 다음과 같다.

① 채권번호와 당첨번호의 마지막 한자리가 같으면 6등상이고 상금 10원을 받는다.

② 채권번호와 당첨번호의 마지막 두자리가 같으면 5등상이고 상금 50원을 받는다.

③ 채권번호와 당첨번호의 마지막 세자리가 같으면 4등상이고 상금 500원을 받는다.

④ 채권번호와 당첨번호의 마지막 네자리가 같으면 3등상이고 상금 5000원을 받는다.

⑤ 채권번호와 당첨번호의 마지막 다섯자리가 같으면 2등상이고 상금 50000원을 받는다.

⑥ 채권번호와 당첨번호가 모두 같은 사람이 1등에 당첨되며 상금은 500000원이다.

또한 매 복권에 대하여 최고액의 상금만 받도록 규정하였다. 매 복권의 평균 당첨금액을 구하시오.

풀기. 를 복권 한장의 당첨금이라고 하면 의 분포렬은 다음과 같다.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | 500000 | 50000 | 5000 | 500 | 50 | 10 | 0 |
|  | 0.000001 | 0.000009 | 0.00009 | 0.0009 | 0.009 | 0.09 | 0.9 |

그러므로 매 복권의 평균소득 =0.5+0.45+0.45+0.45+0.45+0.9+0=3.2이다.

이것은 당첨할 때 수입금 500만원가운데서 320만원을 당첨자들에게 상금으로 돌려주고 나머지를 기타 관리운영비용으로 쓸수 있다는것을 의미한다.

이 실례에서 알수 있다싶이 복권의 당첨여부는 우연적으로 결정되지만 그 복권의 평균소득을 미리 계산해낼수 있으며 그것은 복권설계의 기초이다.

실례 2.2.4 가 구간 우에서 평등분포에 따른다고 가정할 때 를 구하시오.

풀기. 실례2.1.7로부터 의 밀도함수를 알수 있다.



그러므로 .

의 값은 구간 에서 평등하게 값을 취하므로 그의 평균값은 자연히 의 "중간점" 즉 가 되여야 한다.

실례2.2.5 코씨분포에 대한 수학적기대값은 존재하지 않는다는것을 설명하여보자. 코씨분포의 밀도함수는 이므로 이고 따라서 는 존재하지 않는다.

### 2.2.3 수학적기대값의 성질

는 우연량 에 대한 수학적기대값의 정의에 따라 그 분포에 의해 유일하게 결정된다. 만일 우연량 의 함수 (여전히 우연량이다)에 대한 수학적기대값을 구하려면 응당 먼저 의 분포를 구하고 그것을 리용하여 를 구해야 한다. 이 과정을 아래의 실례를 들어 설명할수 있다.

실례2.2.6 우연량 의 분포렬은 아래와 같다.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | -2 | -1 | 0 | 1 | 2 |
|  | 0.2 | 0.1 | 0.1 | 0.3 | 0.3 |

의 수학적기대값을 구하려면 두 단계로 나누어 진행해야 한다.

첫째, 먼저 의 분포를 구한다. 이것은 의 분포로부터 알수 있다.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | (-2)2 | (-1) 2 | 02 | 12 | 22 |
|  | 0.2 | 0.1 | 0.1 | 0.3 | 0.3 |

다음 같은 값을 가지는것들을 합치고 대응되는 확률을 더하면

|  |  |  |  |
| --- | --- | --- | --- |
|  | 0 | 1 | 4 |
|  | 0.1 | 0.4 | 0.5 |

둘째, 의 분포를 리용하여 를 구한다. 즉 .

만일 우리가 같은 값들을 합치기 전의 분포를 리용하여 를 구해도 같은 결과를 얻을수 있다.

.

이 두 방법은 본질적으로 같지만 둘째 방법의 계산은 실제로는 의 분포 (-2, 0, 1, 2)에 기초하여 진행되며 취하는 값을 ((-2)2, (-1)2, 02, 12, 22)로 바꾸어 계산된다. 더 나아가서 를 구할 때 먼저 과 의 분포를 구할 필요가 없으며 아래와 같이 직접 의 분포를 리용하여 구할수 있다.

.

.

일반적인 경우에 다음과 같은 정리를 리용할수 있다.

정리 2.2.1 우연량 의 분포가 분포렬  또는 밀도함수 로 표시된다면 의 함수 에 대한 수학적기대값은 다음과 같다.



여기서 언급된 수학적기대값들은 모두 존재성을 가정한다.

이 정리의 증명은 다른 도구를 필요로 하므로 여기서 생략하였다. 정리 2.2.1을 기초로 수학적기대값의 몇가지 일반적인 성질을 증명하는데 아래에서는 모두 관련된 수학적기대값이 존재한다고 가정한다.

성질 2.2.1 가 상수이면 다.

증명. 상수 를 하나의 값만을 취하는 우연량 로 본다면 이며 따라서 수학적기대값은 이다.

성질 2.2.2 임의의 상수 에 대하여



증명. 식 (2.2.3)에서 라고 놓고 를 합기호 또는 적분기호로부터 내보내면 얻을수 있다.

성질 2.2.3 임의의 두 함수 와 에 대하여



증명. 식 (2.2.3)에서 라고 놓은 다음 합을 두개의 합으로 나누거나 적분을 두개의 적분으로 나누면 증명할수 있다.

실례 2.2.7 어떤 회사가 일종의 원료를 판매하는데 과거 자료에 따르면 그 원료의 시장 수요량 (단위:톤)는 (300, 500)우에서의 평등분포에 따르며 회사는 그 원료 1톤을 판매할 때마다 1500원의 리익을 얻을수 있고 1톤을 사장시키면 500원의 손실을 보게 된다. 회사는 원천을 얼마나 준비해야 평균수익을 최대화할수 있는가?

풀기. 회사가 그 원천을 톤으로 준비한다고 하면 분명히 이다. 또 를 톤의 상품원천이 주어진 조건하에서 수익액(단위:1000원)으로 표시하면 는 수요량 의 함수 즉 이다. 일 때 톤의 상품원천은 모두 판매되며 총 1.5의 리익을 얻는다. 일 때 톤이 판매(리익은 1.5)되고 또 톤이 사장(리익은 )되여 전체 리익은 이다. 이로부터



정리 2.2.1과 평등분포라는 사실로부터



우의 계산으로부터 는 의 2차함수이며 극값을 구하는 일반적인 방법을 리용하면 톤일 때 가 최대값을 취한다는것을 알수 있다. 즉 회사는 450톤의 공급원천을 준비하여야 한다.

련습문제 2.2

1. 리산우연량 의 분포렬이 다음과 같다.

|  |  |  |  |
| --- | --- | --- | --- |
|  | -2 | 0 | 2 |
|  | 0.4 | 0.3 | 0.3 |

이때 와 를 구하시오.

2. 어떤 상점은 지난 몇년간의 판매자료에 근거하여 한명의 손님이 상점에서 소비한 금액 (단위:100원)의 분포를 다음과 같이 정하였다.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | 0 | 1 | 2 | 3 | 4 | 5 |
|  | 0.10 | 0.33 | 0.31 | 0.13 | 0.09 | 0.04 |

이때 손님이 상점에서 소비하는 평균금액을 구하시오.

3. 어떤 지역에서 한달안에 발생한 교통사고건수 는 다음의 분포에 따른다.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|  | 0.301 | 0.362 | 0.216 | 0.087 | 0.026 | 0.006 | 0.002 |

이 지역에서 발생하는 교통사고의 월평균건수를 구하시오.

4. 어떤 화물선의 갑판에 원료가 들어있는 20개의 통이 놓여있는데 그중 5통이 바다물에 오염되였다. 임의로 8통을 선택하고 그가운데서 오염된 통의 수를 라고 하면 의 분포렬과 를 구하시오.

5. 천평으로 어떤 물품의 질량을 측정한다. 기준추는 한 쟁반에만 놓을수 있다. 지금 3개 조의 기준추 (단위:g)가 있는데 (a): 1, 2, 2, 5, 10, (b): 1, 2, 3, 4, 10, (c): 1, 1, 2, 5, 10이다. 무게를 잴 때에는 한조의 기준추만 리용할수 있다. 물품의 질량이 1g, 2g, …, 10g일 확률이 같으면 어느 조의 기준추를 리용하여 무게를 측정하여야 리용하는 평균기준추의 수가 가장 적은가?

6. 같은 종류의 전기부속품이 10개 있고 그중 불합격품이 2개 있다고 가정하자. 계기를 조립할 때에는 이 부속품들가운데서 어느 하나를 가져오는데 그것이 불합격품일 때에는 버리고 다시 어느 하나를 가져온다. 그래도 불합격품이면 버리고 또 하나를 택한다. 합격품을 꺼내기전까지 꺼낸 불합격품수의 수학적기대값을 구하시오.

7. 한 더미의 제품에 대해 검사를 진행한다. 만일 째 제품까지 모두 합격품이면 이 제품더미는 합격품으로 인정하고 째 제품을 검사하기전에 불합격품이 발견되면 검사를 중지하고 이 제품더미는 불합격품으로 인정한다. 제품이 매우 많고 매번 불합격품을 발견할 확률은 라고 가정하면 매 더미의 제품에 대하여 평균 몇개를 검사해야 하는가?

8. 어떤 학생이 1번 문제와 2번 문제 두개가 있는 알아맞추기에 참가하였는데 이 두 문제에 대답하는 순서는 그 자신이 결정할수 있다. 만일 그가 먼저 문제 에 대답하면 답이 옳을 때에만 다른 문제를 대답할수 있다. 문제1을 맞힐 확률이 60%이고 1을 맞힌다면 200원의 상금을 받을수 있다. 또 80%의 확률로 2번 문제를 맞힐수 있으며 2번 문제를 맞히면 100원의 상금을 받게 된다. 어떤 질문에 먼저 대답해야 받는 상금의 수학적기대값이 가장 커지는가?

9. 어떤 기업가가 1만원으로 어떤 주식에 투자하려고 하는데 해당 주식의 현재 가격이 개당 2원이고 1년후 개당 1원과 4원이 될 가능성은 같다고 가정한다. 재정고문은 그에게 1년후 최대시가를 기대한다면 지금 주식을 사고 1년후 보유한 주식의 수량이 최대로 될것을 바란다면 1년후에 살것을 권고한다. 이 재정고문의 조언이 맞는가? 왜 그런가?

10. 어느 한 보험회사의 규정에 의하면 발생할 확률이 인 어떤 사건 가 1년안에 발생하였다면 보험회사는 보험가입자에게 원을 지불해야 한다. 보험회사가 보험가입자로부터 받는 보험료가 일 경우 가 얼마여야 보험회사수익의 기대값이 의 10%까지 되게 할수 있는가?

11. 어느 한 공장에서 고장난 뜨락또르의 수리시간 는 하나의 우연량(단위: 시간)으로서 그 밀도함수는 다음과 같다.



평균수리시간을 구하시오.

12. 새로운 제품이 앞으로의 시장에서 차지할 점유률 는 구간 (0, 1)에서 값을 취하는 우연량이며 그 밀도함수는 이다. 시장점유률의 평균을 구하시오.

13. 우연량 의 밀도함수가 다음과 같을 때 를 구하시오.



14. 우연량 의 분포함수가 다음과 같을 때 를 구하시오.



15. 우연량 의 밀도함수가 다음과 같다.



만일 이면 와 를 구하시오.

16. 어떤 작업이 끝나는 시간 (단위:월)는 우연량이며 그 분포렬은 다음과 같다.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | 10 | 11 | 12 | 13 |
|  | 0.4 | 0.3 | 0.2 | 0.1 |

① 이 작업을 완성하는 평균월수를 구하시오.

② 이 작업으로 얻는 리윤이 (단위:만원)일 때 평균리윤을 구하시오.

③ 계획을 바꾸어 작업을 완성하는 시간 (단위:월)의 분포가 다음과 같을 때 평균리윤이 얼마나 증가할수 있는가?

|  |  |  |  |
| --- | --- | --- | --- |
|  | 10 | 11 | 12 |
|  | 0.5 | 0.4 | 0.1 |

17. 우연량 의 확률밀도함수가 다음과 같다.



에 대하여 독립적으로 4번 반복하여 관측하였을 때 가 관측값이 보다 큰 관측차수를 표시할 때 에 대한 수학적기대값을 구하시오.

18. 우연량 의 밀도함수를 다음과 같이 가정한다.



이때 의 수학적기대값을 구하시오.

19. 를 부아닌 옹근수만 취하는 리산우연량이라고 하고 그의 수학적기대값이 존재한다고 하면 다음과 같은 관계가 성립한다는것을 증명하시오.

① .

② .

20. 련속우연량 의 분포함수가 이고 수학적기대값이 존재한다고 할 때 다음을 증명하시오.



21. 를 비부인 련속우연량이라고 하고 가 존재하면 다음을 증명하시오.

① .

② .

## 2.3 우연량의 분산과 표준편차

우연량 에 대한 수학적기대값 는 분포의 위치특징수로서 의 값이 의 주위에서 변동한다는것을 반영한다. 그러나 위치특징수는 우연량이 취하는 값의 "변동크기"를 나타내지 못한다. 례를 들어 와 의 분포는 각각 다음과 같다.

|  |  |  |  |
| --- | --- | --- | --- |
|  | -1 | 0 | 1 |
|  | 1/3 | 1/3 | 1/3 |

|  |  |  |  |
| --- | --- | --- | --- |
|  | 10 | 0 | 10 |
|  | 1/3 | 1/3 | 1/3 |

이것들의 수학적기대값이 모두 0이지만 값의 변동은 값의 변동보다 크다. 우연량이 취하는 값이 "변동"하는 크기를 어떻게 수값으로 반영할것를 이 절에서 론의한다. 아래에서 정의하는 분산과 표준편차는 변동을 측정하는 가장 중요한 특징수이다.

### 2.3.1 분산과 표준편차의 정의

우연량 의 평균을 라고 하면 의 값이 꼭 인것은 아니며 편차가 있을수 있다. 편차량 는 정일수도 있고 부일수도 있다. 정과 부인 편차량들이 서로 상쇄되지 않도록 하기 위하여 일반적으로를 고려하며 수학적으로 처리하기 어려운 절대값 을 고려하지 않는다. 는 또 다른 우연량이기때문에 그 평균값 는 의 "변동"정도를 나타낼수 있으며 이것을 의 분산이라고 한다. 그 정의는 다음과 같다.

정의 2.3.1 우연량 의 수학적기대값 가 존재한다면 두제곱편차 의 수학적기대값 를 우연량 (또는 그에 대응하는 분포)의 **분산**이라고 하고 다음과 같이 표시한다.



그리고 분산의 두제곱뿌리 를 우연량 (또는 그에 대응하는 분포)의 **표준편차**라고 하며  또는 로 표시한다.

분산과 표준편차는 그 기능이 비슷한데 모두 우연량의 값이 집중, 분산되는 정도를 나타내는 두개의 특징수이다. 분산과 표준편차가 작을수록 우연량의 값은 더 집중되고 반대로 클수록 우연량의 값이 더 분산된다.

분산과 표준편차의 차이점은 주로 단위에 있다. 왜냐하면 표준편차는 론의되는 우연량, 수학적기대값과 같은 단위를 가지기때문에 (여기서 는 정의 실수) 역시 의미가 있다. 실천적으로 표준편차를 더 많이 리용하지만 표준편차의 계산은 분산을 통하여서만 가능하다.

주의할것은 우연량 에 대한 수학적기대값이 존재하여도 그 분산이 반드시 존재하는것은 아니다. 그러나 의 분산이 존재할 때 는 반드시 존재하는데 그 리유는 이 항상 성립하기때문이다.

실례 2.3.1 그림 2.3.1에 3각형분포, 평등분포, 거꿀3각형분포의 밀도함수와 그것들의 그라프가 있다. 그림으로부터 이 세 분포는 모두 구간 (-1, 1)우에 있고 세로축대칭이므로 기대값은 모두 0이라는것을 알수 있다. 그러나 그것들의 분산은 다르고 따라서 표준편차도 다르다. 이것은 매개의 분포에 의해 계산될수 있다.



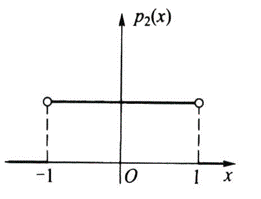
이러한 계산결과는 분포에 대한 직관적인 인식과 일치한다.이 세 분포에서 삼각분포의 확률은 중간에 비교적 집중되여있으므로 분산이 가장 작다. 거꿀3각형분포의 확률은 대부분 량측에 분산되여있으므로 분산이 가장 크다. 세 분포의 수학적기대값, 분산, 표준편차를 그림 2.3.1에서 해당 분포의 오른쪽에 표시하였다.

① 3각형분포





② 평등분포





③ 거꿀3각형분포

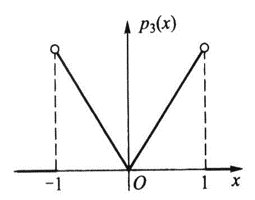




그림 2.3.1 세개 분포의 분산과 표준편차의 비교

실례 2.3.2 어떤 기업가가 부동산과 상업에 투자할수 있는 자금을 가지고있으며 그 수익은 모두 시장상태와 관련된다. 앞으로의 시장을 상, 중, 하의 3등급으로 나누었는데 발생 확률은 각각 0.2, 0.7, 0.1이다. 조사를 통하여 얻은 이 기업가가 부동산에 투자한 수익 (만원)과 상업에 투자한 수익 (만원)의 분포는 각각 다음과 같다.

|  |  |  |  |
| --- | --- | --- | --- |
|  | 11 | 3 | -3 |
|  | 0.2 | 0.7 | 0.1 |

|  |  |  |  |
| --- | --- | --- | --- |
|  | 6 | 4 | -1 |
|  | 0.2 | 0.7 | 0.1 |

그렇다면 기업가는 어떻게 투자하는것이 좋은가?

풀기. 먼저 수학적기대값(평균수익)을 보자.

(만원)

(만원)

평균수익으로 보면 부동산에 투자하는 수익이 더 크다. 이제 매개의 분산을 계산해보자.

,

.

표준편차는 이다.

표준편차나 분산이 클수록 수익의 변동이 크고 따라서 위험도 크다. 자금을 어디에 투자하는지는 평균수익이 얼마인지를 보아야 할뿐만 아니라 위험의 크기도 고려해야 한다. 여기서 표준편차로 보면 부동산에 투자하는 위험이 상업에 투자하는 위험보다 배나 더 크다. 수익과 위험을 종합적으로 따져보면 그래도 상업투자를 선택하는것이 좋다. 비록 평균수익이 1000원이 적지만 위험은 절반이상 적다.

### 2.3.2 분산의 성질

아래에서는 우연량의 분산이 존재한다고 가정한다.

성질 2.3.1 .

증명. 이므로 수학적기대값의 성질 2.2.3으로부터

.

분산의 실제 계산에서 이 성질은 정의인 보다 더 잘 리용된다.

성질 2.3.2 상수의 분산은 0 즉 이다. 여기서 는 상수이다.

증명. 가 상수이면 .

성질 2.3.3 가 상수이면 이다.

증명. 가 상수이기때문에 .

한편 이므로 이면 이고 또 이다.

실례 2.3.3 를 주사위를 한개 던질 때 나타나는 점수로 볼 때 를 구하시오.

풀기.



### 2.3.3 체븨쉐브부등식

아래에 확률론에서 중요한 기본부등식을 제시한다.

정리 2.3.1(체븨쉐브(Chebyshev,1821-1894)부등식) 우연량 에 대한 수학적기대값과 분산이 존재한다면 임의의 상수 에 대해서 다음의 부등식이 성립한다.



또는



증명. 가 련속우연량이고 그 밀도함수는 라고 하자. 라고 표시하면



이로부터 식 (2.3.2)이 련속우연량에 대해 성립한다는것을 알수 있으며 리산우연량에 대해서도 류사하게 증명할수 있다.

확률론에서 사건 을 **큰 편차**라고 하며 그 확률 를 **큰 편차발생확률**이라고 한다. 체븨쉐브부등식은 큰 편차가 발생할 확률의 상계를 주었는데 이 상계는 분산에 정비례하며 분산이 클수록 상계도 크다.

다음의 정리는 분산이 0이면 우연량이 취하는 값들이 거의 한 점에 집중되여있음을 의미한다는것을 설명한다.

정리 2.3.2 우연량 의 분산이 존재한다면 의 필요충분조건은 가 거의 모든 곳에서 어떤 상수 인것 즉 인것이다.

증명. 충분성은 명백하므로 아래에서 필요성만 증명한다. 라고 가정하면 가 존재한다. 또한 이므로



여기서 마지막 부등식은 체븨쉐브부등식을 리용하여 얻었다. 따라서 이며 결국 . 즉 . 이렇게 결론을 증명하였으며 상수 는 바로 이다.

련습문제 2.3

1. 우연량 가 를 만족하고 이 주어졌을 때 를 구하시오.

2. 같은 종류의 전기부속품이 10개 있고 그중 불합격품이 2개라고 가정한다. 계기를 조립할 때에는 이 부속품들가운데서 어느 하나를 가져오는데 그것이 불합격품일 때에는 버리고 다시 어느 하나를 가져온다. 그래도 불합격품이면 버리고 또 하나를 택한다. 합격품을 꺼내기전까지 꺼낸 불합격품수의 분산을 구하시오.

3. 이 주어졌을 때 를 구하시오.

4. 라고 하자. 만일 이면 을 구하시오.

5. 우연량 의 분포함수가 다음과 같다.



이때 를 구하시오.

6. 우연량 의 밀도함수가 다음과 같다.



이때 을 구하시오.

7. 우연량 의 밀도함수가 다음과 같다.



라고 하고 를 구하시오.

8. 우연량 의 분포함수가 일 때 와 를 구하시오.

9. 임의의 상수 에 대하여 이라는것을 증명하시오.

10. 우연량  가 구간 에서 값을 취한다고 할 때 를 증명하시오.

11. 우연량 의 값 을 취할 확률은 각각 이다. 이라는것을 증명하시오.

12. 는 우연량 의 값모임에서 정의된 부가 아닌 비감소함수이고 가 존재한다고 할때 임의의 에 대하여 라는것을 증명하시오.

13. 를 부가 아닌 우연량, 이라고 할 때 가 존재한다면 임의의 에 대하여 라는것을 증명하시오.

14. 정상인 남성어른의 혈액에서 백혈구수는 리터당 평균 이며 표준편차는 이다. 체븨쉐브부등식을 리용하여 혈액 리터당 백혈구수가 부터 사이에 있을 확률의 하한을 추정하시오.

## 2.4 대표적인 리산분포

우연량마다 하나의 분포가 있으며 서로 다른 우연량은 분포가 같을수도 있고 다를수도 있다. 우연량은 수천만개이지만 자주 리용하는 분포는 그리 많지 않다. 이런 자주 리용하는 분포에 익숙하면 다른 분포를 쉽게 인식할수 있다. 흔히 쓰는 분포는 리산분포와 련속분포의 두 종류로 나뉜다. 이 절에서는 리산분포를 론하고 다음 절에서 련속분포를 론한다.

### 2.4.1 2항분포

1. 2항분포

만일 를 중베르누이시행에서 성공(사건 로 표시한다)한 차수로 본다면 의 가능한 값은 0, 1, …, 이다. 를 매번의 실험에서 가 발생할 확률 즉 이라고 하면 이다.

중베르누이시행의 기본결과를 와 같이 표시할수 있다. 여기서 는  또는 일수 있는데 이런 가 모두 개 있다. 이 개의 표본점 들이 표본공간 를 구성한다.

이제 의 분포렬 즉 사건 의 확률을 구한다. 어떤 표본점에 대하여 이라는것은 가운데 개는 이고 개는 이라는것을 의미한다. 따라서 독립성으로부터 임을 알수 있다. 또한 사건 에는 이런 가 개 있으므로 의 분포렬은

.

이 분포를 **2항분포**라고 하고 로 표시한다.

그 합이 1이라는것을 쉽게 알수 있다. 즉 .

이로부터 2항확률 는 바로 차2항식 의 전개식에서 째 항이라는것을 알수 있다. 이것이 바로 그 이름의 유래이다.

2항분포는 자주 리용되는 리산분포의 한 종류이다. 실례를 들면

• 10개 제품을 검사하는데 그중 불합격품의 개수 는 2항분포 에 따르며 여기서 는 불합격률이다.

• 50명을 조사하는데 그중 색맹에 걸린 사람의 수 는 2항분포 에 따르며 여기서 는 색맹률이다.

• 5번 사격하는데 명중차수 는 2항분포 에 따르며 는 사격수의 명중률이다.

실례 2.4.1 어떤 약의 림상유효률은 0.95이다. 현재 10명이 복용하고있을 때 적어도 8명이 완치될 확률은 얼마인가?

풀기. 를 10명중 완치된 사람의 수라고 하면 이고 구하려는 확률은 다음과 같다.



즉 10명중 적어도 8명이 완치될 확률은 0.9884이다.

실례2.4.2 우연량 , 라고 하자. 일 때 를 구하시오.

풀기. 이므로 . 따라서 이고 이며 로부터 임을 알수 있다.

2. 2점분포

일 때의 2항분포 을 **2점분포** 또는 **0-1분포**, **베르누이분포**라고 한다. 그 분포렬은



또는

|  |  |  |
| --- | --- | --- |
|  | 0 | 1 |
|  |  |  |

2점분포 는 주로 베르누이시행에서 성공(로 표시)한 회수(0 또는 1)를 설명하는데 리용된다.

많은 우연현상의 표본공간은 와 의 두 부분으로 나누며 이렇게 베르누이시행을 형성한다. 중베르누이시행은 독립적으로 수행되는 개의 동일한 베르누이시행으로 이루어진다. 번째 베르누이시행에서 가 나타난 회수를 로 표시한다면 중베르누이시행에서 매 베르누이시행은 서로 독립이므로 개의 우연량 은 서로 독립이고 같은 2점분포 에 따른다. 그리고 그 합 은 중베르누이시행에서 가 나타난 총차수인데 이것은 2항분포 에 따른다. 이것이 2항분포 와 2점분포 사이의 관계이다. 즉 2항분포에 따르는 우연량은 같은 2점분포에 속하는 독립적인 개 우연량의 합으로 분해할수 있다.

3. 2항분포의 수학적기대값과 분산

우연량 에 대하여



또한



따라서 의 분산은 다음과 같다.

.

2점분포는 일 때의 2항분포 이기때문에 2점분포의 수학적기대값은 이고 분산은 이다.

서로 다른 에 따르는 2항분포 의 변화정형을 알아보기 위하여 표 2.4.1에 일 때의 2항분포확률값을 보여주었으며 그라프를 그림 2.4.1에 주었다.

표 2.4.1 일부 2항분포의 확률값(빈칸은 0)

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|  | 0.107 | 0.268 | 0.302. | 0.201 | 0.088 | 0.027 | 0.006 | 0.001 |  |  |  |
|  | 0.001 | 0.010 | 0.044 | 0.117 | 0.205 | 0.246 | 0.205 | 0.117 | 0.044 | 0.010 | 0.001 |
|  |  |  |  | 0.001 | 0.006 | 0.027 | 0.088 | 0.201 | 0.302 | 0.268 | 0.107 |

(a) 의 선도표(우측편기)

(b) 의 선도표(대칭)

(c) 의 선도표(좌측편기)

그림 2.4.1 2항분포 의 선도표

0.3

0.2

0.1

0

0

1

2

3

4

5

6

7

1

2

3

4

5

6

7

8

4

5

6

7

8

9

10

11

9

우의 그림으로부터 다음과 같은것을 알수 있다.

• 평균값 의 주위에서 확률이 크다.

• 가 증가함에 따라 분포의 봉우리는 점차 오른쪽으로 이동한다.

실례 2.4.3 두 장기수 A, B는 10판을 두어 이긴 판수가 많은 쪽이 이긴다고 약속하였다. 매 판에서 A, B가 이길 확률은 각각 0.6, 0.4이다. 매 판이 독립적으로 진행된다면 A가 이기는 확률, B가 이기는 확률, 승부가 나지 않을 확률은 각각 얼마인가?

풀기. 10판의 경기에서 A가 이긴 판수를 로 표시하면 이다. 그러므로 A가 이기는 사건 , B가 이기는 사건 , 승부가 나지 않는 사건 의 확률은 각각 다음과 같다.



이렇게 A가 이길 가능성은 63.3%이지만 B가 이길 가능성은 16.63%밖에 안되는데 이것은 승부가 나지 않을 가능성보다 더 작다. 마지막의 두 확률을 더한 합 0.3670은 B가 지지 않을 확률을 나타낸다.

### 2.4.2 뽜쏭분포

1. 뽜쏭분포

뽜쏭분포는 1837년 프랑스의 수학자 뽜쏭(Poisson, 1781-1840)에 의해 처음 제안되였다. 뽜쏭분포의 확률분포렬은 다음과 같다.

.

여기서 파라메터 이며 로 표시한다.

뽜쏭분포의 경우 그 합이 1이라는것을 쉽게 확인할수 있다.

.

뽜쏭분포는 일반적으로 단위시간(또는 단위면적, 단위제품 등)에 대한 계수과정과 련관된 리산분포의 한가지로써 실례로

• 하루동안 어떤 상점을 찾은 손님수

• 단위시간내에 한 전기회로가 외부전자기파의 충격을 받는 회수

• 1㎡의 유리우에 있는 기포의 개수

• 일정한 시간내에 어떤 방사성물질에서 방출되는 -립자수

등은 모두 뽜쏭분포에 따른다. 따라서 뽜쏭분포는 매우 널리 리용된다.

2. 뽜쏭분포의 수학적기대값과 분산

우연량 에 대하여 .

이것은 뽜쏭분포 에 대한 수학적기대값이 파라메터 라는것을 의미한다.

또한





이므로 의 분산은 이다.

즉 뽜쏭분포 에서 파라메터는 수학적기대값이면서 분산이기도 하다.

서로 다른 에 따르는 뽜쏭분포 의 변화정형을 알아보기 위하여 표 2.4.2에 0.8, 2.0, 4.0일 때의 뽜쏭분포확률값을 보여주었으며 그라프를 그림 2.4.2에 주었다.

표 2.4.2 일부 뽜쏭분포의 확률값(빈칸은 0)

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|  | 0.449 | 0.360 | 0.144 | 0.038 | 0.008 | 0.001 |  |  |  |  |  |
|  | 0.135 | 0.271 | 0. 271 | 0.180 | 0.090 | 0.036 | 0.012 | 0.004 | 0.001 |  |  |
|  | 0.018 | 0.073 | 0.147 | 0.195 | 0.195 | 0.156 | 0.104 | 0.060 | 0.030 | 0.013 | 0.005 |

(a) 의 선도표

(b) 의 선도표

(c) 의 선도표

그림 2.4.2 뽜쏭분포 의 선도표

0.3

0.2

0.1

0

0

1

2

3

4

5

0

1

2

3

4

5

6

7

2

3

4

5

6

7

8

9

0.4

0

1

우의 그림으로부터 다음과 같은것을 알수 있다.

• 평균값 의 주위에서 확률이 비교적 크다.

• 가 증가하는데 따라 분포가 점차 대칭화된다.

실례2.4.4 한 주물품에서 기포(불량)의 수는 인 뽜쏭분포에 따른다. 기포가 최대 1개일 확률과 최소 2개일 확률을 구하시오.

풀기. 가 이 주물품의 기포수라고 할 때 이다. 따라서 기포가 최대로 1개 있을 확률은 다음과 같다.

.

최소 2개의 기포가 있을 확률은 이다.

실례 2.4.5 어떤 상점에서 판매되는 상품의 지난 기간의 판매기록을 분석한데 의하면 월판매량은은 파라메터가 8인 뽜쏭분포에 따른다. 월초에 입고할 때 재고량이 얼마나 있어야 손님들의 수요를 90%이상 만족시킬수 있겠는가.

풀기. 를 이 상품의 월판매량이라고 하면 이다. 그러면 수요를 만족시키는것은 이 성립되는 최소의 정의 옹근수 이다.

이런 을 찾기 위하여 부록 1에 있는 여러 에 대한 뽜쏭분포 의 수표를 리용한다. 일 때 부록 1로부터 이므로 월초에 12개의 상품을 입고하면 90%이상의 가능성으로 손님들의 수요를 만족시킬수 있다.

3. 2항분포의 뽜쏭근사

뽜쏭분포의 또 다른 매우 실용적인 성질은 뽜쏭분포를 2항분포에 대한 근사로 볼수 있다는것이다. 2항분포 에서 이 비교적 클 때 계산량은 대단히 크다. 그러나 이 크고 가 작을 때 다음과 같은 뽜쏭정리를 리용하면 2항분포의 계산량을 줄일수 있다.

정리 2.4.1(뽜쏭정리) 중베르누이시행에서 사건 가 발생할 확률은 으로서 실험회수 과 련관이 있다. 만일 일 때 이면



증명. 라고 표시하면 이므로



고정된 에 대하여





.

따라서 은 임의의 에 대하여 성립한다. 이렇게 정리가 증명된다.

뽜쏭정리는 인 조건하에서 얻었으므로 2항분포 를 계산할 때 이 매우 크고 는 작아서 그 적 의 크기가 적당할 때 뽜쏭분포를 근사로 리용할수 있다. 즉

.

표 2.4.3은 2항분포에 따르는 직접적인 계산과 뽜쏭분포로 근사한 구체적인 수값을 보여준다. 표에서 알수 있는바와 같이 두 결과는 매우 비슷하며 이 클수록, 가 작을수록 근사정도가 좋다.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 2항분포 | | | | 뽜쏭근사 |
|  | | | |  |
|  |  |  |  |  |
| 0 | 0.349 | 0.358 | 0.363 | 0.366 | 0.368 |
| 1 | 0.387 | 0.377 | 0.373 | 0.370 | 0.368 |
| 2 | 0.194 | 0.189 | 0.186 | 0.185 | 0.184 |
| 3 | 0.057 | 0.060 | 0.060 | 0.061 | 0.061 |
| 4 | 0.011 | 0.013 | 0.014 | 0.015 | 0.015 |
| >4 | 0.002 | 0.003 | 0.004 | 0.003 | 0.004 |

아래에서 뽜쏭분포를 리용한 근사계산의 몇가지 실례를 준다.

실례 2.4.6 어떤 질병의 발병률이 0.001이고 어느 단위에 모두 5000명이 있다. 그러면 그 단위에서 이런 질병에 걸릴 사람이 5명을 초과하지 않을 확률은 얼마인가?

풀기. 그 단위에서 이런 질병을 앓는 사람의 수를 라고 하면 이다. 그러면 구하려는 확률은 이다. 이 계산량은 대단히 많다. 이 크고 가 작으므로 뽜쏭분포를 리용하여 근사시키면 이다.

실례 2.4.7년령과 직업이 비슷한 10000명이 손전화기보험회사의 보험에 가입하였다. 매 보험가입자에게는 오직 한대의 손전화기만 있고 매해 년초에 200원의 보험료를 납부해야 하며 이 1년안에 손전화기가 파손되는 경우 가입자는 보험회사로부터 10만원의 배상비를 받을수 있다. 통계에 의하면 손전화기의 파손률은 0.001로 알려져있다. 보험회사의 이 업무에서 다음과 같은 확률을 구하시오.

① 손해볼 확률

② 적어도 50만원의 리득을 볼 확률

풀기. 를 10000명의 보험가입자가운데서 1년간 파손된 손전화기대수라고 하면 는 2항분포 에 따른다. 이 업무에서 1년간 보험회사의 총수입은 원이다. 이 크고 은 매우 작으므로 인 뽜쏭분포를 리용하여 근사계산할수 있다.

① 보험회사가 이 업무에서 손해본다는것은 사건 이 발생한것과 같으며 따라서 구하려는 확률은 다음과 같다.

.

이로부터 보험회사가 이 업무에서 손해볼 가능성은 매우 작다는것을 알수 있다.

② 보험회사가 이 업무에서 최소 50만원의 리득을 본다는것은 사건 가 발생한것과 같다. 따라서 구하려는 확률은 다음과 같다.

.

이로부터 보험회사가 이 업무에서 적어도 50만원의 리익을 볼 가능성이 매우 크다는것을 알수 있다.

실례 2.4.8 설비의 정상적인 작업을 보장하기 위해서는 수리공을 배치해야 한다. 만일 매 설비가 고장나는 사건은 서로 독립이고 고장날 확률이 모두 0.01이라고 하면 아래의 매 경우에 설비가 고장이 생기지만 제때에 수리하지 못할 확률을 구하시오.

① 1명의 수리공이 20대의 설비를 책임진다.

② 3명의 수리공이 90대의 설비를 책임진다.

③ 10명의 수리공이 500대의 설비를 책임진다.

풀기. ① 로 20대의 설비가운데서 동시에 고장이 발생하는 대수를 표시하면 이다. 파라메터가 인 뽜쏭분포를 리용하여 근사계산하면 확률은 다음과 같다.

.

② 90대의 설비가운데서 동시에 고장이 발생한 대수를 로 표시하면 이다. 인 뽜쏭분포를 리용하여 근사계산하면 확률은 다음과 같다.

.

주의할것은 이 경우의 확률이 첫 경우보다 작아졌을뿐아니라 3명의 수리공가 90대의 설비를 책임지면 한 수리공이 30대의 설비를 책임지는것과 같으며 작업효률은 첫 경우의 1.5배라는것이다.

③ 500대의 설비가운데서 동시에 고장이 발생한 대수를 으로 표시하면 이다. 파라메터 인 뽜쏭분포로 근사계산하면 확률은 다음과 같다.

.

이 경우 확률은 둘째 경우와 거의 같지만 수리공 10명이 500대의 설비를 책임지는것은 매 수리공이 50대를 책임지는것과 같으며 작업효률은 둘째 경우의 1.67배, 첫 경우의 2.5배이다.

이로부터 수리공들이 공동으로 많은 설비를 책임지면 작업의 효률을 높일수 있다는것을 알수 있다.

### 2.4.3 초기하분포

1. 초기하분포

초기하분포는 유한모집단으로부터 비반환표본화를 할 때 맞다들린다.

개의 제품중에서 개가 불합격품이라고 하자. 만일 비반환방식으로 개를 우연적으로 발취하면 그중에 포함된 불합격품의 수 는 초기하분포에 따르며 로 표시한다. 초기하분포의 확률분포렬은

.

여기서 이고 이며 은 모두 정의 옹근수이다(실례 1.2.3을 참고하시오).

우의 식이 확률분포렬이라는것을 조합항등식(련습문제 1.2) 을 리용하면 쉽게 검증할수 있다.

초기하분포는 널리 리용되는 리산분포의 일종으로서 표본화리론에서 중요한 위치를 차지한다.

2. 초기하분포의 수학적기대값과 분산

라면 에 대한 수학적기대값은 다음과 같다.

.

또한



이므로 의 분산은 이다.

3. 초기하분포의 2항근사

일 때 매번 발취한 후 모집단에서의 불합격품률 은 거의 변하지 않으므로 비반환표본화를 반환표본화로 근사하게 볼수 있다. 이때 초기하분포는 2항분포로 근사시킬수 있다.



여기서 이다.

2.4.4기하분포와 부의 2항분포

1. 기하분포

베르누이시행렬의 매 시행에서 사건 가 발생할 확률을 로 표시하자. 만일 를 사건 가 처음 발생하였을때의 시행회수라면 의 가능한 값은 1, 2, …이다. 이때 는 기하분포에 따른다고 하는데 로 표시한다. 그 분포렬은 다음과 같다.

.

실제 문제에서 많은 우연량들이 기하분포에 따르는데 실례를 들면

• 한 제품의 불합격률이 0.05이면 처음 불합격품을 발견할 때까지의 검사회수 

• 한 사수의 명중률이 0.8이면 처음 목표를 맞힐 때까지의 사격회수 

• 어떤 주사위를 던질 때 처음 6점이 나올 때까지 던진 회수 

• 동시에 2개의 주사위를 던질 때 두개 점수의 합이 처음으로 8일 때까지 던진 회수 

2. 기하분포의 수학적기대값과 분산

우연량 가 기하분포 에 따른다고 하자. 라고 놓으면 항별 미분을 통하여 의 수학적기대값을 얻을수 있다.



또한



이므로 의 분산은 다음과 같다.

.

기하분포의 수학적기대값으로부터 알수 있는것처럼 주사위를 던질 때 처음으로 6점이 나오는 평균던지기회수는 6번이다.

3. 기하분포의 무기억성

정리 2.4.2(기하분포의 무기억성) 라고 하면 임의의 정의 옹근수 에 대하여

.

증명에 앞서 먼저 우의 확률같기식의 의미를 해석한다. 하나의 베르누이시행렬에서 만일 첫 성공 ()이 출현하는 시행회수 가 기하분포에 따른다면 사건 은 처음 번의 시행에서 가 발생하지 않았다는것을 의미한다. 만일 다음 번의 시행에서 가 여전히 발생하지 않으면 이 사건을 으로 표시한다. 이 정리는 앞서 번의 시행에서 가 발생하지 않는 조건에서 다음 번의 시행에서 가 여전히 나타나지 않을 확률은 과만 관계될뿐 이전의 번의 시행과는 관계가 없다는것을 설명하여준다. 이것은 마치 앞선 번의 시행결과를 잊어버린것 같다는것인데 이것이 바로 무기억성이다.

증명.

이므로 임의의 정의 옹근수 에 대하여 조건부확률

.

이렇게 식 (2.4.9)를 증명한다.

4. 부의 2항분포

기하분포의 일반화로서 **부의 2항분포**(**파스칼분포**라고도 부른다)를 론의한다. 베르누이시행렬의 매 시행에서 사건 가 발생할 확률은 이라고 하자. 만일 를 사건 가 번 발생할 때까지의 시행회수라고 하면 의 가능한 값은 이다. 이때 는 부의 2항분포 또는 파스칼분포에 따른다고 한다. 그 분포렬은



이며 로 표시한다. 이면 기하분포로 된다.

그것은 번의 베르누이시행에서 마지막 한번은 반드시 이고 앞의 번에서는 가 번 발생해야 하기때문이다. 또한 2항분포로부터 그 확률이 이고 다시 마지막으로 가 발생할 확률 를 곱하면 식 (2.4.10)을 얻을수 있다.

부의 2항분포의 수학적기대값은 이고 분산은 이다. 처음 가 출현하는 평균시행수는 이고 번째로 가 발생하는 평균시행수는 이기때문에 이것은 직관적으로 보아도 부합된다.

만일 가 처음 출현하는 시행회수를 , 두번째로 출현하는 시행회수(첫 가 출현한 후부터 계산)를 , …, 번째로 출현하는 시행회수(째 가 출현한 후부터 계산)를 로 표시(그림 2.4.3)하면 들은 독립동일분포이고 이다. 이때  즉 부의 2항분포에 따르는 우연량을 개의 독립동일분포에 따르는 우연량들의 합으로 분해할수 있다.



그림 2.4.3 부의 2항분포우연량과 기하분포우연량사이의 관계

례를 들어 많은 제품을 련속적으로 검사할 때 그 제품의 불합격률이 0.05이면 검사에서 다섯째 불합격품이 발견될 때의 검사회수 는 부의 2항분포 에 따르며 그 평균검사회수는 이다.

자주 리용되는 리산분포들을 표 2.5.1에 주었다.

련습문제 2.4

1. 한 무지의 제품가운데 10%의 불합격품이 있을 경우 그중 3개를 임의로 선택할 때 기껏 한개가불합격품일 확률을 구하시오.

2. 한 자동화된 생산선에서 생산된 제품의 합격품률은 0.8이다. 5개를 검사할 때 적어도 2개의 합격품이 있을 확률을 구하시오.

3. 어떤 사수가 10점을 맞힐 확률은 0.7, 9점을 맞힐 확률은 0.3이다. 세번 사격할 때 29점이상 맞힐 확률을 구하시오.

4. 식당에 좌석을 예약해놓고 오지 않는 손님의 비률이 20%이라고 하자. 현재 식당에 50개의 좌석이 있지만 52명이 예약했다. 이때 손님이 식당에 와서 좌석이 없을 확률은 얼마인가?

5. 라고 하자. 일 때 파라미터 를 구하시오.

6. 우연량  가 2항분포 에 따르고 우연량 가 2항분포 에 따른다고 하자. 이면 는 얼마인가?

7. 어떤 제품의 불합격률이 0.02일 때 제품더미에서 40개를 임의로 선택하여 검사를 진행한다. 만일 2개이상의 불합격품이 발견되면이 더미의 제품을 불통과한다. 아래와 같은 방법으로 불통과될 확률을 구하시오.

① 2항분포를 리용한 정확한 계산

② 뽜쏭분포를 리용한 근사계산

8. 가 뽜쏭분포에 따르고 이라면 를 구하시오.

9. 하루동안 어떤 상점에 오는 손님수 는 파라메터가 인 뽜쏭분포에 따르고 손님들이 상점에 와서 물건을 사는 확률은 이다. 이때 이 상점에서 하루동안 물건을 구입하는 손님수는 파라메터가 인 뽜쏭분포에 따른다는것을 증명하시오.

10. 한 사람이 1년동안 감기에 걸리는 회수는 인 뽜쏭분포에 따른다고 하자. 어떤 감기예방약은 75%의 사람들에게게 효과(뽜쏭분포의 파라메터를 으로 줄일수 있다)가 있지만 25%에게는 효과가 없다. 만일 어떤 사람이 이 약을 먹고 1년동안 두번 감기에 걸렸다면 그 약이 그 사람에게 효과가 있을 가능성은 얼마나 되는가?

11. 세 친구가 커피를 마시러 갔는데 그들은 동전을 던지는 방식으로 누가 돈을 낼것인지를 결정하였다. 매 사람이 동전을 던지는데 한사람이 던진 결과가 다른 두 사람과 다르면 그가 돈을 낸다. 만일 세사람이 던진 결과가 같다면 돈을 낼 사람이 결정될 때까지 계속 다시 던진다. 다음과 같은 사건의 확률을 구하시오.

① 2회전까지 던졌을 때 누가 돈을 낼것인지 확정하였다.

② 3회전까지 던졌지만 아직 돈낼 사람을 확정하지 못하였다.

12. 개의 흰 공, 개의 검은 공이 들어있는 주머니에서 반환을 하면서 흰공이 나올 때까지 공을 꺼내본다. 검은 공을 꺼낸 개수의 수학적기대값을 구하시오.

13. 어떤 제품에서 흠집의 개수 는 다음과 같은 분포렬에 따른다.

.

이 제품의 평균흠집수를 구하시오.

14. 우연량 의 밀도함수가 다음과 같다.



에 대한 세차례의 독립적인 반복관찰에서 사건 이 나타나는 회수를 로 표시할 때 를 구하시오.

15. 한 제품의 불합격률은 0.1인데 매번 10개를 우연적으로 선택하여 검사한다. 불합격품이 1개를 넘으면 설비를 조정한다. 검사원이 하루에 4번씩 검사를 한다면 설비를 하루 평균 몇번씩 조정해야 하는가?

16. 하나의 회로는 여러개의 부속품으로 구성되였으며 매 부속품이 정상적으로 동작하는가 하는것은 서로 독립이며 또 매 부속품이 정상적으로 동작하는 확률은 이다. 만일 회로에서 적어도 절반의 부속품들이 정상동작하면 전체 회로는 유효하다. 가 어떤 값을 취할 때 5개 소자로 된 회로가 3개 소자로 된 회로보다 더 유효하겠는가?

17. 우연량 가 파라메터 인 뽜쏭분포에 따른다고 할 때 를 증명하고 이것을 리용하여 를 계산하시오.

18. 2항분포 에 따르는 우연량 에 대하여 임을 증명하시오.

19. 우연량 가 파라메터가 인 기하분포에 따른다고 할 때 을 증명하시오.

20. 우연량 에 대하여 을 증명하시오.

## 2.5 대표적인 련속분포

련속분포의 경우 밀도함수와 분포함수는 서로 유도해낼수 있고 같은 정보를 포함하며 각기 다른 용도가 있다. 그러나 그라프의 측면에서 밀도함수는 련속분포의 여러가지 특성을 직관적으로 나타낼수 있다. 례를 들어 정상상태, 치우침상태, 단봉우리와 평평한 봉우리 등은 모두 밀도함수의 그라프에서 론의된것이기때문에 사람들은 더욱 밀도함수에 더 관심을 가진다.

### 2.5.1 정규분포

정규분포는 확률론과 수리통계에서 가장 중요한 분포이며 가우스(Gauss, 1777-1855)가 오차리론을 연구할 때 처음으로 정규분포를 리용하여 오차의 분포를 서술하였다. 그러기때문에 정규분포를 가우스분포라고도 한다.

이 책의 4장에서 중심극한정리는 한 우연량이 많고 작으며 또 독립인 우연요인들이 겹쳐 작용하여 얻어진 결과라면 이 우연량은 정규분포에 따른다고 볼수 있다는것을 설명하여준다. 그러므로 많은 우연량을 정규분포로 묘사하거나 근사적으로 묘사할수 있다. 실례로 측정오차, 제품 중량, 사람의 키, 년 강수량 등을 정규분포를 리용하여 묘사할수있다.

1. 정규분포의 밀도함수와 분포함수

우연량 의 밀도함수가

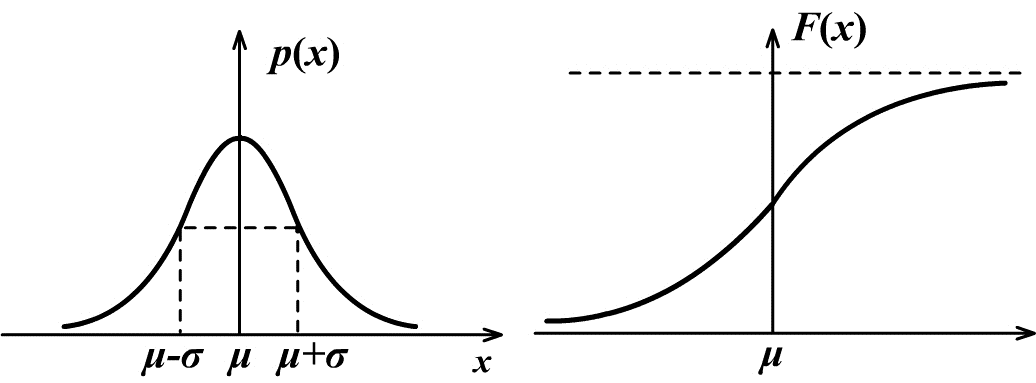


이면 를 **정규분포**에 따른다고 하며 를 **정규우연량**이라고 하며 로 표시한다. 이 경우 파라메터는 이다. 밀도함수 의 밀도함수의 그라프를 그림 2.5.1(a)에 표시하였다. 는 방울(bell)형곡선으로서 중간은 높고 량쪽은 낮으며 에 관하여 좌우대칭이다. 는 정규분포의 중심이며 주위에서 값을 취할 가능성이 크고 량쪽에서 값이 취할 가능성은 적다. 는 이 곡선의 변곡점이다.

(a) 밀도함수 

(b) 분포함수 

그림 2.5.1 정규분포



정규분포 의 함수는 다음과 같다.



이것은 그림 2.5.1(b)와 같이 미끈하게 증가하는 형곡선이다.

그림 2.5.2는 와 가 변할 때 대응하는 정규분포밀도곡선의 변화를 보여준다.

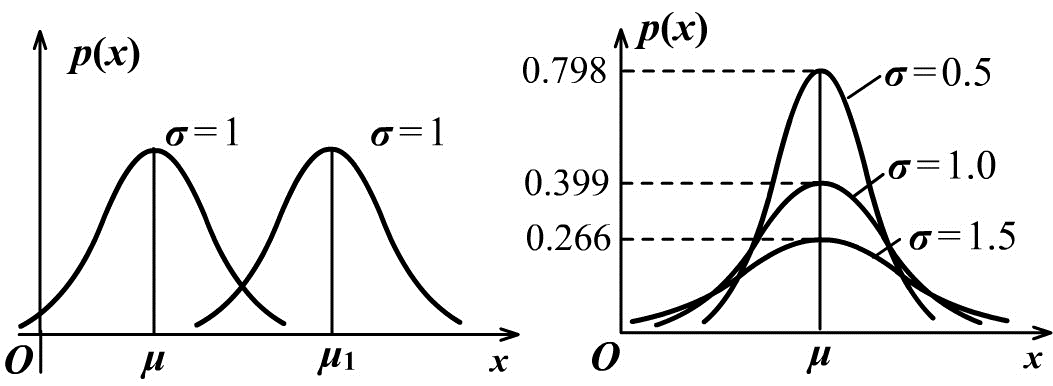
• 그림 2.5.2(a)에서 볼수 있듯이 를 고정하고 의 값을 변화시키면 그라프의 모양은 변하지 않고 축을 따라 이동한다. 즉 정규분포밀도함수의 위치는 파라메터 에 의해 결정된다. 따라서 를 **위치파라메터**라고 한다.

• 그림 2.5.2(b)에서 볼수 있듯이 를 고정하고 의 값을 변화시키면 분포의 위치는 변하지 않지만 가 작을수록 곡선이 높고 가늘며 비교적 집중되여있다. 가 클수록 곡선은 낮고 두터우며 분산되여있다. 즉 정규분포밀도함수의 척도는 파라메터 에 의해 결정된다. 따라서 를 **척도파라메터**라고 한다.

(a)  고정,  변화

(b)  고정,  변화

그림 2.5.2 정규분포밀도함수



2. 표준정규분포

일 때의 정규분포 을 **표준정규분포**라고 한다.

보통 표준정규분포우연량을 로, 표준정규분포의 밀도함수를 로, 분포함수를 로 표시한다. 즉



표준정규분포의 분포함수는 아무런 미지파라메터도 포함하지 않기때문에 그 값 는 완전히 계산할수 있다. 부록 2에 에 대한 의 값을 주었다. 에 대하여 다음과 같은 식이 성립한다는것을 쉽게 증명할수 있다.

• 

• 

• 

• 

실례 2.5.1 일때 부록 2를 리용하여 다음 사건들의 확률을 구하시오.

① .

② .

③ .

④



⑤ .

3. 정규분포우연량의 표준화

정규분포들의 족 에서 표준정규분포 은 그 족의 중심성원이다. 다음의 정리는 일반적으로 정규분포우연량은 어떤 선형변환(표준화)를 통하여 표준정규분포우연향으로 변환할수 있다는것을 설명한다. 따라서 정규우연량과 관련된 모든 사건의 확률은 표준정규분포함수표를 찾아 구할수 있다. 이로부터 표준정규분포 은 일반정규분포 의 계산에서 중요한 역할을 한다.

정리 2.5.1 우연량 에 대하여 이 성립한다.

증명. 와 의 분포함수를 각각 라고 하면 분포함수의 정의로부터

.

정규분포함수는 엄격한 단조증가함수이고 임의의 점에서 미분가능하므로 와 의 밀도함수를 각각 , 로 표시하면 이다.

이로부터 이다.

우의 정리로부터 실천에서 널리 쓰이는 계산공식들을 얻을수 있다. 이제 이면

.



실례 2.5.2 우연량 가 정규분포 에 따른다고 할 때 다음을 구하시오.

① .

② 인 상수 .

풀기. 식 (2.5.4)과 부록 2를 리용한다.

①



② 이므로 인데 여기서 은 의 거꿀함수이다. 부록 2로부터 이고 다시 선형보간법을 리용하면 이다. 즉 이며 이다. 따라서 이다.

우의 실례에서 알수 있는것처럼 어떤 경우에는 의 값 가 주어지면 부록 2를 리용하여  또는 가 만족하는 를 구할수 있다. 이때 를 표준정규분포의 분위수라고 한다.우의 례에서 1.645는 표준정규분포의 0.95분위수이며 더 일반적인 설명은 2.7.3절에 주었다. 분위수도 통계학에서 많이 리용되는 개념이다.

실례 2.5.3시험에서 응시자의 점수 가 근사하게 정규분포에 따른다면 이 시험은 일반적으로 응시자의 성적등급을 합리적으로 분류하는 측면에서 정상이라고 본다. 교원들은 보통 점수가 를 넘으면 A급, 부터 까지는 B급, 부터 까지는 C급, 부터 까지는 D급, 이하이면 E급으로 본다. 이로부터 다음과 같이 계산된다.







.

따라서 이와 같은 방법으로 등급을 나눈 경우 A급은 약 16%, B급은 약 34%, C급은 약 34%, D급은 약 14%, E급은 약 2%라는것을 알수 있다.

4. 정규분포의 수학적기대값과 분산

우연량 에 대하여 이므로 의 수학적기대값은 이다. 여기서 피적분함수는 홀함수이기때문에 그 적분값은 0 즉 이다. 또한 이므로 수학적기대값의 성질로부터 이다. 즉 정규분포 에서의 는 수학적기대값이다.

그리고



이고 이므로 분산의 성질로부터 이다.

이것은 정규분포 의 다른 파라메터 은 의 분산이라는것을 말해준다.

정규분포의 수학적기대값과 분산을 구하는 과정에서 변환 이 리용된다. 즉 . 많은 경우에 이 변환을 리용할수 있는데 특히 비정규분포우연량 에 대해서도 리용할수 있다. 만일 의 수학적기대값 와 분산 이 존재하면 을 의 **표준화우연량**이라고 한다. 이때 이다.

5. 정규분포의 법칙

우연량 에 대하여 이다.

일 때



이것은 정규분포의 중요한 성질로서 어떤 우연량이 값을 취할 확률이 대체적으로 식 (2.5.5)을 만족하면 이 우연량은 근사하게 정규분포에 따른다고 할수 있다. 그러나 식 (2.5.5)의 세개가운데서 하나의 편차가 큰 경우 이 우연량은 정규분포에 따르지 않는다고 볼수 있다. 이것이 정규분포의 법칙이며 이것을 리용하여 의 관측값이 비교적 많을 때(수백개 또는 수천개) 의 분포가 정규분포에 근사한가를 판단할수 있다.

어떤 제품의 생산과정에 제품의 품질에 대한 아래웃한계를 규정하는데 만일 이 구간이 을 피복하면 이 생산공정이 제한받는다고 하며 그 비률 을 생산공정능력지수라고 한다. 여기서 은 각각 웃, 아래 한계이다. 일 때 그 생산공정은 불비한것으로 보며 일 때 정상이라고 본다. 가 기타 다른 값을 취할 때에는 생산공정이 불안정하다고 보고 개선하여야 한다.

### 2.5.2 평등분포

1. 평등분포의 밀도함수와 분포함수

앞에서 실례를 들어 평등분포를 설명하였지만 여기서는 일반적인 설명을 준다. 우연량 의 밀도함수(그림 2.5.3(a))가



이면 는 구간 우에서 **평등분포**에 따른다고 하며 로 표시한다. 그 분포함수(그림 2.5.3(b))는



(a) 밀도함수 

(b) 분포함수 

그림 2.5.3 구간 우의 평등분포























평등분포는 실천에서 자주 리용된다. 례를 들면 반경 인 자동차다이야 원둘레우의 임의의 점이 땅면과 접촉할 가능성은 꼭 같으므로 땅면과의 접촉위치 는 우에서의 평등분포에 따른다. 이것은 페기된 자동차다이야의 마모된 정도가 모든 위치에서 거의 같은것을 보면 잘 알수 있다.

실례2.5.4 우연량 가 우의 평등분포에 따른다고 하자. 에 대하여 네번의 독립적인 관측을 진행할 때 적어도 세번의 관측값이 5보다 클 확률을 구하시오.

풀기. 네번의 독립적인 관측에서 관측값이 5보다 큰 회수를 라고 하면 이다. 여기서 이다. 이므로 의 밀도함수는



그러므로 이고 따라서

.

2. 평등분포의 수학적기대값과 분산

우연량 에 대하여 . 이것은 구간 의 가운데점이다.

그리고 이므로 의 분산은

.

실례로 평등분포 의 평균값은 이고 분산은 이다.

### 2.5.3 지수분포

1. 지수분포의 밀도함수와 분포함수

우연량 에 대한 밀도함수(그림 2.5.4)가

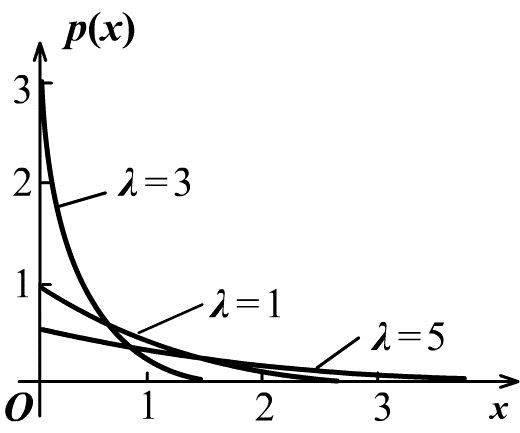


이면 는 **지수분포**에 따른다고 말하며 로 표시한다. 여기서 파라메터 이다.

지수분포의 분포함수는 다음과 같다.



그림 2.5.4 파라메터가 인 지수분포밀도함수



지수분포는 일종의 편향된 분포이며 우연량이 부가 아닌 실수만을 취할수 있기때문에 일반적으로 전자제품의 수명, 동물의 수명, 전화통화시간, 우연봉사체계에서의 대기시간 등과 같은 여러가지 "수명"의 분포에 리용된다. 지수분포는 믿음성과 대기렬리론에서 널리 리용된다.

2. 지수분포의 수학적기대값과 분산

우연량 에 대하여

.

지수분포에서 로 표시하는 경우가 있는데 이 경우 는 지수분포의 수학적기대값이다.

그리고 이므로 의 분산은 다음과 같다.

.

례를 들어 전자 부속품의 수명(단위:시간) 는 인 지수분포 에 따르는데 이때 평균수명은 1000(시간), 분산은 (시간)이다. 수명자료의 분산은 일반적으로 매우 크다.

3. 지수분포의 무기억성

리산분포의 경우 정리 2.4.2에서 기하분포의 무기억성을 제시하였다면 련속분포의 경우 다음의 정리에서 지수분포의 무기억성을 제시한다.

정리 2.5.2(지수분포의 무기억성) 우연량 가 있을 때 임의의 에 대하여

.

다시 말하여 가 어떤 제품의 사용수명(시간)일 때 가 지수분포에 따른다면 이미 시간을 고장없이 리용하였다는것을 아는 상태에서 다시 시간 고장없이 리용할 확률은 이미 리용한 시간 와 아무런 관계가 없으며 오직 시간동안 다시 리용하는 확률과 같다는것을 의미한다. 즉 이미 리용한 시간에 대한 기억이 없다.

증명. 이므로 이고 또 이므로 조건부확률은

.

이렇게 식 (2.5.10)이 증명되였다.

다음 실례들은 뽜쏭분포와 지수분포의 관계를 보여준다.

실례 2.5.5 어떤 설비가 길이가 인 시간 에 고장나는 회수 (시간의 길이 와 관련)가 파라메터가 인 뽜쏭분포에 따른다면 두번의 련속적인 고장사이의 시간간격 는 파라메터가 인 지수분포에 따른다는것을 설명하시오.

증명. 라고 하면 이다.

시간간격 는 부아닌 우연량이며 사건 은 이 설비가 에서 고장나지 않았다는것을 의미한다. 즉 . 이로부터 일 때 이며 일 때 이다. 따라서 이다. 그림 2.5.5는 이 관계를 보여준다.



고장회수 

0





고장간격 

그림 2.5.5 고장회수와 고장간격사이의 관계

### 2.5.4 감마분포

1. 감마함수

다음과 같은 함수



를 **감마함수**라고 하는데 여기서 파라메터 이다. 감마함수는 다음과 같은 성질을 가진다.

① .

② . 또한 가 자연수 일 때 .

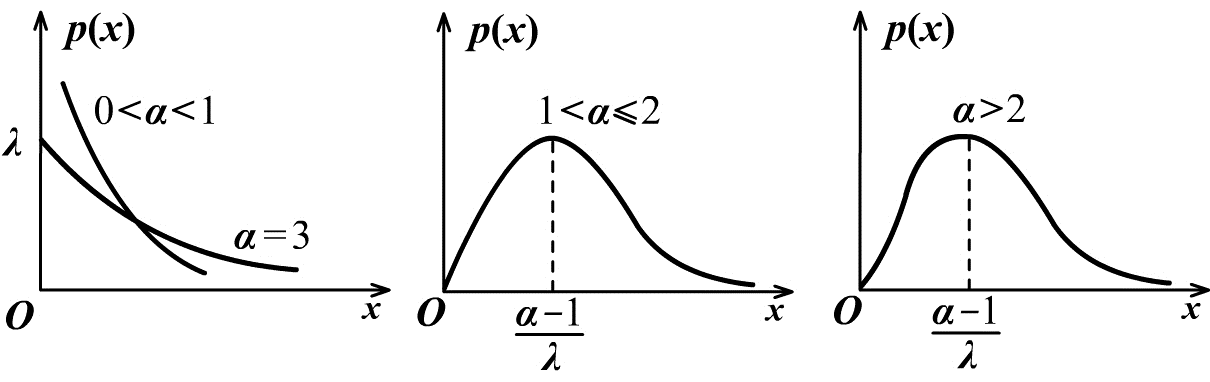
2. 감마분포

우연량 의 밀도함수가 다음과 같다고 하자.



이때 는 **감마분포**에 따른다고 하며 로 표시한다. 여기서 은 형태파라메터, 은 척도파라메터이다. 그림 2.5.6에 가 고정되고 가 서로 다른 감마분포의 밀도함수곡선을 보여주고 있다. 그림으로부터 다음과 같은것을 알수 있다.

그림 2.5.6 가 고정되고 가 서로 다른 감마분포밀도함수곡선



(a)

(b)

(c)

• 일 때 는 엄격한 감소함수이며 이 특이점이다.

• 일 때 는 엄격한 감소함수이며 에서 이다.

• 일 때 는 단봉우리함수이며 먼저 우로불룩이고 후에 아래우묵이다.

• 일 때 는 단봉우리함수이며 먼저 아래불룩이고 중간에서 우로불룩이며 후에 아래불룩이다. 가 클수록 는 정규밀도함수에 근사하지만 감마분포는 항상 편기된 분포이며 가 작을수록 그 편기정도가 심하다.

3. 감마분포 의 수학적기대값과 분산

감마함수의 성질을 리용하여 감마분포 에 대한 수학적기대값을 쉽게 계산할수 있다.

.

그리고 이므로 의 분산은 다음과 같다.

.

4. 감마분포의 두가지 특수경우

감마분포에는 일반적으로 리용되는 두가지 특수경우가 있다.

① 일 때 감마분포는 지수분포이다. 즉



② 일 때의 감마분포는 자유도가 인 **분포**이다. 이것을 로 표시한다. 즉

.

밀도함수는 다음과 같다.



여기서 은 분포의 유일한 파라메터로서 자유도라고 하며 정의 실수를 취할수도 있지만 대부분 정의 옹근수를 취한다. 분포는 통계응용에서 하나의 중요한 분포이다.

분포가 특수한 감마분포이므로 감마분포의 기대값과 분산을 통하여 분포의 기대값과 분산을 쉽게 구할수 있다.

.

실례 2.5.6 전자제품은 외부적인 "충격"에 의해 고장나는 경우가 많다. 만일 내에 발생하는 충격의 차수 가 를 파라메터로 하는 뽜쏭분포에 따른다면 번째 충격을 받는 시간 은 감마분포 에 따른다는것을 증명하시오.

증명. 사건 "번째 충격을 받는 시간 이 보다 작거나 같다"는 사건 "내에 충격이 발생하는회수 가 보다 크거나 같다"와 동등하다. 즉 . 따라서 의 분포함수는 다음과 같다.

. 부분적분법을 리용하면

.

그래서 이고 가 증명된다.

### 2.5.5 베타분포

1. 베타함수

다음과 같은 함수



를 **베타함수**라고 하며 여기서 는 파라메터이다.

베타함수는 다음과 같은 성질을 가진다.

① .

증명. 식 (2.5.17)의 적분에서 라고 놓으면

.

② 베타함수와 감마함수사이에는 다음과 같은 관계가 있다.



증명. 감마함수의 정의로부터 이다. 변수변환 를 진행하고 그 야꼬비행렬식 이므로



이렇게 식 (2.5.18)을 증명한다.

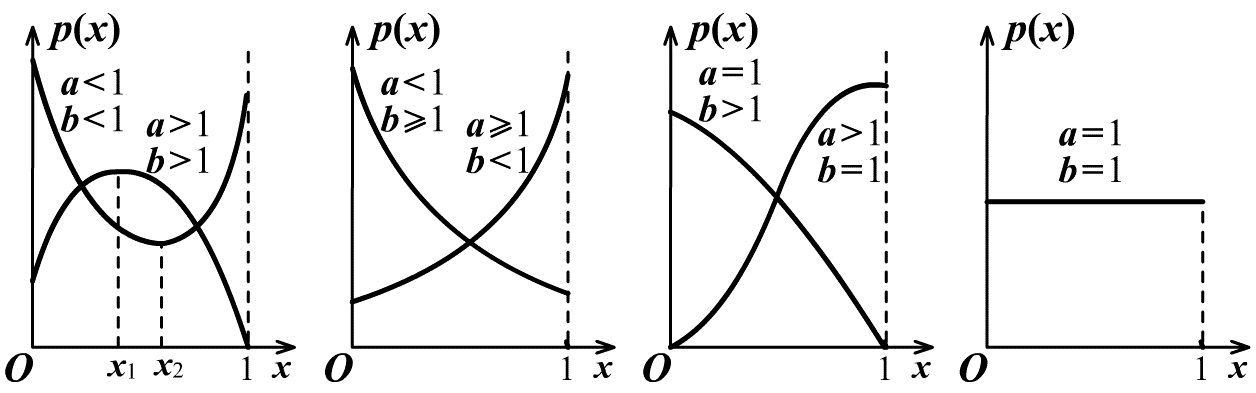
2. 베타분포

우연량 의 밀도함수가 다음과 같다고 하자.



이때 는 **베타분포**에 따른다고 하며 로 표시한다. 여기서 은 모두 형태파라메터이다.그림 2.5.7에 대표적인 베타분포의 밀도함수곡선을 보여주었다.

그림 2.5.7 베타분포의 밀도함수곡선



그림으로부터 다음과 같은것을 알수 있다.

• 일 때 는 아래불룩한 형함수이다.

• 일 때 는 우로불룩한 단봉우리함수이다.

• 일 때 는 아래불록한 단조감소함수이다.

• 일 때 는 아래볼록한 단조증가함수이다.

• 일 때 이다.

베타분포 에 따르는 우연량은 구간 에서만 값을 취하기때문에 불합격품률, 기계의 정비률, 시장점유률, 사격의 명중률 등 여러가지 비률에 대하여 베타분포를 확률분포로 리용할수 있다. 이때 파라메터 를 적절히 선택하면 된다.

3. 베타분포 의 수학적기대값과 분산

베타함수의 성질을 리용하여 베타분포 의 수학적기대값을 계산할수 있다.

.

또한



이므로 의 분산은 다음과 같다.

.

일반적인 분포와 그 수학적기대값, 분산을 표 2.5.1에 보여주었다.

표 2.5.1일반적으로 리용되는 확률분포와 그 수학적기대값, 분산

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 분포 | 분포렬  또는 분포밀도 | | 수학적기대값 | 분산 |
| 0-1분포 |  | |  |  |
| 2항분포 |  | |  |  |
| 뽜쏭분포 |  | |  |  |
| 초기하분포 |  |  |  |  |
| 기하분포 |  | |  |  |
| 부의 2항분포 |  | |  |  |
| 정규분포 |  | |  |  |
| 평등분포 |  | |  |  |
| 지수분포 |  | |  |  |
| 감마분포 |  | |  |  |
| 분포 |  | |  |  |
| 베타분포 |  | |  |  |
| 로그정규분포 |  | |  |  |
| 꼬씨분포 |  | | 존재안함 | 존재안함 |
| 웨이블분포 |  | |  |  |

주의: 표에는 매 분포에 대한 밀도함수의 비령구역만 렬거하였다.

련습문제 2.5

1. 우연량 가 구간 (2, 5)우에서 평등분포에 따른다고 할 때 에 대한 세번의 독립적인 관측에서 적어도 2번의 관측값이 3보다 클 확률을 구하시오.

2. (0, 1)에서 임의로 취한 한 점을 로 표시할 때 를 구하시오.

3. 가 (1, 6)우의 평등분포에 따른다고 할 때 방정식 이 실뿌리를 가질 확률을 구하시오.

4. 우연량 에 대하여 방정식 이 실뿌리를 가지지 않을 확률이 0.5 라고 할 때 를 구하시오.

5. 2의 저항을 흐르는 전류의 세기 가 9~11A범위에 있을 때 이 저항에서 소비된 평균전력을 구하시오. 여기서 전력은 이다.

6. 어떤 원판의 직경은 구간 우에서 평등분포하는데 이런 원반들의 평균면적을 구하시오.

7. 어떤 상품의 매주 수요량 가 구간 (10, 30)우에서 평등분포하고 상점에 입고되는 량은 구간 (10, 30)의 어떤 옹근수라고 하며 상품 1개를 판매할 때마다 500원의 리윤을 얻는다고 하자. 만일 공급이 수요를 초과하면 가격을 인하하여 처리하는데 상품 1개당 100원의 손해를 보게 된다. 또한 공급이 수요를 보장하지 못할 경우 다른 상점으로부터 보충받을수 있는데 이때 상품 1개당 리윤이 300원밖에 안된다. 상점에서 얻는 평균리윤이 9280원보다 적지 않게 하려면 입고량을 최소로 얼마로 확정하여야 하는가?

8. 통계조사에 의하면 1875년부터 1951년까지 잉글랜드의 광산에서 10명 또는 그 이상이 사망한 련이은 사고의 시간간격 (일)는 평균값이 241인 지수분포에 따른다. 를 구하시오.

9. 한번의 전화통화시간 (분)은 파라메터가 0.25인 지수분포에 따른다. 평균통화시간을 구하시오.

10. 어떤 설비의 사용수명 (년)는 지수분포에 따르며 그 평균수명은 4년이다. 이 설비를 만든 공장에서는 리용한지 1년내에 설비가 파손되면 교환해준다고 규정하고 있다. 만일 공장이 설비 한대를 팔 때마다 100원의 리윤을 보고 설비 한대를 교체하는데 300원이 든다면 설비 한대의 평균리윤을 구하시오.

11. 손님이 은행에서 봉사를 기다리는 시간 (분)가 다음과 같은 지수분포에 따른다고 하자.



어떤 손님이 봉사를 대기하고 있다가 10분이 지나면 자리를 뜬다. 그는 1년에 5번 은행에 간다. 1년동안 그가 봉사를 기다리지 않고 떠나가는 회수를 로 표시하면 를 구하시오.

12. 어떤 계기에 독립적으로 작업하는 같은 형의 전자부속품 3개가 조립되여있는데 그 수명 (시간)는 모두 동일한 밀도함수가 다음과 같은 지수분포에 따른다.



이 계기가 처음 리용되는 200시간내에 적어도 하나의 전자부속품이 못쓰게 될 확률을 구하시오.

13. 우연량 의 밀도함수가 다음과 같다.



이때 가 되는 를 구하시오.

14. 우연량 의 밀도함수가 다음과 같다.



이때 인 의 값범위를 구하시오.

15. 다음 정규분포들의 평균값과 표준편차를 구하시오.

.

16. 어떤 지역에서 사는 18살 청년들의 혈압 (수축기압, mmHg단위)는 에 따른다. 혈압이 100~120일 가능성이 얼마인가를 구하시오.

17. 어떤 지역에서 사는 성인남자의 체중 (kg단위)는 정규분포 에 따른다. , 이 주어졌을 때 다음의 값들을 구하시오.

① 와 .

② 우연적으로 5명의 성인남성을 선택하였을 때 그중 적어도 2명이 65 kg이상일 확률.

18. 어떤 기계로 만든 볼트의 길이(cm단위)는 정규분포 에 따른다. 길이가 (10.050.12)cm범위에 있으면 합격품이라고 할 때 볼트가 불합격될 확률을 구하시오.

19. 한 지역에 대한 표본조사에 의하면 수험생의 외국어성적(100점 기준)은 근사하게 인 정규분포에 따른다. 수험생수의 2.3% 가 96점이상을 받은 경우 수험생의 성적이 60점이상일 확률을 구하시오.

20. 우연량 에 대하여

① 를 구하시오.

② 를 구하시오.

③ 인 를 결정하시오.

21. 우연량 에 대하여

① 를 구하시오.

② 를 구하시오.

③ 이 만족되는 와 그의 최대값을 구하시오.

22. 어떤 목표물까지의 거리를 측정할 때 발생하는 우연오차 (m단위)는 다음과 같은 밀도함수를 가진다.



세번의 측정에서 적어도 한번 오차의 절대값이 30m를 넘지 않을 확률을 구하시오.

23. 지점 A에서 B로 가는 렬차는 매일 오전 10시 10분에 출발하며 주행시간 의 평균값은 4시간, 표준편차는 20분이다.

① 렬차가 오후 2시 30분이후에 B지점에 도착할 확률은 얼마인가?

② 렬차가 오후 2시 20분이전에 B지점에 도착할 확률은 얼마인가?

③ 렬차가 오후 1시 50분~2시 30분사이에 B지점에 도착할 확률은 얼마인가?

24. 어떤 대학입학시험에 모두 1만명이 응시하였다. 시험성적이 정규분포에 따른다고 하고 90점이상을 받은 사람은 359명, 60점이하를 받은 사람은 1151명이라는것을 안다고 하자. 시험성적이 높은 사람으로부터 순서에 따라 2500명을 입학시킨다면 그들중 가장 낮은 점수는 몇점인가?

25. 우연량 에 대하여 가 다음과 같은 5개 구간에 놓일 확률의 비가 7:24:38:24:7로 되는 를 구하시오.



26. 우연량 와 가 각각 정규분포 에 따를 때 의 크기를 비교해보시오.

27. 우연량 에 대하여 이라면 를 구하시오.

28. 우연량 에 대하여 의 증가에 따라 확률 는 어떻게 변화하는가?

29. 우연량 가 파라메터 인 정규분포에 따른다고 할 때 이 만족되는 의 최대값은 얼마인가?

30. 우연량 에 대하여 를 구하시오.

31. 우연량 에 대하여 를 증명하시오.

32. 우연량 에 대하여 를 구하시오.

33. 어떤 학급의 학생들가운데서 수학성적이 불합격되는 비률 는 인 베타분포에 따른다. 를 구하시오.

## 2.6 우연량함수의 분포

가 에서 정의된 함수이고 가 우연량이라고 가정하면 는 의 함수로서 역시 우연량이다. 실천문제에서 흥미를 가지는 문제는 우연량 의 분포를 알고있을 때 어떻게 다른 우연량 의 분포를 구하는가 하는것이다.

우연량함수의 분포를 찾는것은 확률론의 기본요령으로서 확률론과 수리통계에서는 자주 리용한다. 아래에서 리산적인 경우와 련속적인 경우에 각각 우연량함수의 분포에 대하여 론의한다.

### 2.6.1 리산우연량함수의 분포

리산우연량함수의 분포는 비교적 쉽게 얻을수 있다. 가 리산우연량이고 그의 분포렬이 다음과 같다고 하자.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  | … |  | … |
|  |  |  | … |  | … |

도 리산우연량이며 이때 의 분포렬은 간단히 다음과 같이 표시할수 있다.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  |  | … |  | … |
|  |  |  | … |  | … |

들가운데서 어떤 값들이 같을 경우 그것들을 통합한 후 대응되는 확률을 더하면 된다.

실례2.6.1 우연량 의 분포렬이 다음과 같을 때 의 분포렬을 구하시오.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | -2 | -1 | 0 | 1 | 2 |
|  | 0.2 | 0.1 | 0.1 | 0.3 | 0.3 |

풀기. 의 분포렬은

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 2 | 0 | 0 | 2 | 6 |
|  | 0.2 | 0.1 | 0.1 | 0.3 | 0.3 |

같은 값을 합치면 다음과 같다.

|  |  |  |  |
| --- | --- | --- | --- |
|  | 0 | 2 | 6 |
|  | 0.2 | 0.5 | 0.3 |

### 2.6.2 련속우연량함수의 분포

리산우연량의 함수는 여전히 리산우연량이다. 그러나 련속우연량 의 함수 는 반드시 련속우연량인것은 아니다. 의 분포를 몇가지 경우로 나누어 론의한다.

1. 가 리산우연량일 때

이 경우에 의 가능한 값을 일일이 렬거한 다음 가 취할수 있는 여러가지 값들의 확률을 구하면 된다. 례를 들어 , 이라고 하면 쉽게 는 인 0-1분포에 따른다는것을 계산해낼수 있다.

2. 가 엄격한 단조함수일 때

이 경우 다음과 같은 정리가 있다.

정리 2.6.1 를 밀도함수가 인 련속우연량이라고 하자. 는 또 다른 련속우연량이다.만일 가 엄격히 단조이고 그 거꿀함수 가 련속인 도함수를 가진다면 의 밀도함수는 다음과 같다.



여기서 이다.

증명. 일반성을 잃지 않고 를 엄격한 단조증가함수로 가정하면 그 거꿀함수 도 엄격한 단조증가함수이며 이다. 라고 표시하면 이것은 가 구간 에서만 값을 취한다는것을 의미한다. 따라서 일 때 , 일 때 , 일 때 이다. 이로부터 의 밀도함수는



마찬가지로 가 엄격한 단조감소감수일 때에도 같은 결론을 얻을수 있다. 그러나 이때 주의해야 할것은 이기때문에 절대값을 취해야 하는데 이때 이다. 이렇게 정리가 증명된다.

우의 정리를 리용하여 몇가지 결론을 정리의 형식으로 표시하고 증명한다.

정리 2.6.2 우연량 가 정규분포 에 따르면 일 때 이다.

증명. 일 때 는 엄격한 증가함수이며 여전히 에서 값을 취한다. 그의 거꿀함수는 이며 정리 2.6.1로부터



이것은 정규분포 의 밀도함수이다.

일 때 는 엄격한 감소함수이며 여전히 에서 값을 취한다. 그의 거꿀함수는 이며 정리 2.6.1로부터 이다.

이것은 정규분포 의 밀도함수이다. 이렇게 정리가 증명된다.

이 정리는 정규우연량의 선형변환은 여전히 정규우연량이며 그의 수학적기대값과 분산은 직접 선형변환으로부터 얻을수 있다는것을 보여준다. 를 취하면 인데 이것은 앞절의 정리 2.5.1이다.

실례 2.6.2

① 우연량 에 대하여 의 분포를 구하시오.

② 우연량 에 대하여 의 분포를 구하시오.

풀기.

① 정리 2.6.2로부터 는 여전히 정규우연량이며 수학적기대값과 분산은 각각 다음과 같다는것을 알수 있다.



그러므로 의 분포는 이다.

② 는 여전히 정규우연량이며 수학적기대값과 분산은 각각 다음과 같다.



따라서 의 분포는 그대로 이다. 이것은 와 의 분포는 같지만 우연량은 다르다는것을 의미한다. 결국 분포가 같다는것과 우연량이 같다는것은 전혀 다른 개념이다.

정리 2.6.3(로그정규분포) 우연량 에 대하여 의 확률밀도함수는



증명. 는 엄격한 증가함수이고 에서만 값을 취하며 그 거꿀함수는 이다. 정리 2.6.1로부터

일 때 이며 따라서 이다.

일 때 의 밀도함수는 이다. (증명끝)

이 분포를 **로그정규분포**라고 부르며 로 표시한다. 여기서 는 로그평균값, 은 로그분산이라고 한다. 로그정규분포는 일반적으로 널리 리용되는 치우친 분포로서 실천에서 다음과 같은 우연량들이 로그정규분포에 따른다.

• 절연재료의 수명

• 고장에 대한 수리시간

• 한 가정에 두 아이가 있을 때 그들의 나이차

정리 2.6.4 우연량 가 감마분포 에 따르고 일 때 이다.

증명. 이므로 는 엄격한 증가함수이고 에서 값을 취하며 그 거꿀함수는 이고 정리 2.6.1로부터 이면 , 이면

.

이것은 의 밀도함수이다. (증명끝)

이 결론은 아주 쓸모가 있는데 례를 들면 일 때  즉 임의의 감마분포는 분포로 변환할수 있다.

정리 2.6.5 우연량 의 분포함수 가 엄격히 단조증가하는 련속함수이고 그 거꿀함수 가 존재하면 ~이다.

증명. 의 분포함수를 구하자. 분포함수 가 [0, 1]구간에서만 값을 취한다. 따라서 일 때 는 불가능사건이고 이다. 일 때 이며 일 때 는 필연사건이므로 이다.

결론적으로 의 분포함수는 다음과 같다.



이것은 평등분포 의 분포함수이므로 이다.

이 정리는 평등분포가 련속분포들가운데서 특수한 지위를 차지한다는것을 말해준다. 임의의 련속우연량 는 그 분포함수 를 통하여 평등분포우연량 와 관계를 맺을수 있다. 실례로 가 지수분포 에 따른다면 그 분포함수는 이고 를 로 바꾸면 , 이다. 따라서 평등분포 의 우연수(허위관측값)으로부터 지수분포 의 우연수 를 얻을수 있다. 그런데 평등분포하는 우연수는 아무런 통계프로그람에서도 생성할수 있으므로 지수분포와 같은 기타분포의 우연수도 얻을수 있게 된다. 이렇게 여러가지 분포의 우연수를 얻는것은 우연모의법(몬테카를로법)의 기초이다.

3. 가 다른 형태일 때

정리 2.6.1을 리용하여 련속우연량 의 분포를 찾기 어려울 때 직접 의 분포함수 로부터 출발하여 함수 의 특성에 따라 구체적인 처리를 진행할수 있다.

실례2.6.3 우연량 가 표준정규분포 에 따를 때 의 분포를 구하시오.

풀기. 먼저 의 분포함수 를 구한다. 이므로 일 때 이고 이다. 이면 이다. 따라서 의 분포함수는 다음과 같다.



다시 도함수를 취하는 방법으로 의 밀도함수를 구하면



분포의 밀도함수와 비교해보면 이라는것을 알수 있다.

실례2.6.4 우연량 의 밀도함수를 다음과 같다.

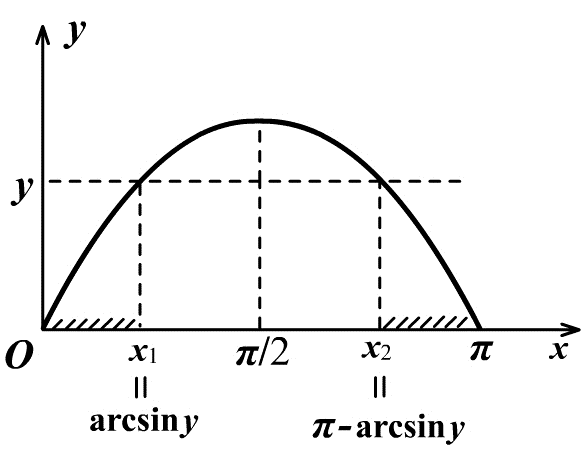


이때 의 밀도함수 를 구하시오..

풀기. 우연량 가 에서 값을 취하므로 의 값범위는 (0, 1]이다. 의 값범위밖에서 이다.

그림 2.6.1에서 볼수 있듯이 일 때 인 의 값들은 서로 겹치지 않는 두개의 구간이다.

그림 2.6.1의 그라프





그리하여



이며 따라서



량변에 대하여 에 관한 도함수를 구하면



련습문제 2.6

1. 리산우연량 의 분포렬이 다음과 같이 주어졌다.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | -2 | -1 | 0 | 1 | 3 |
|  |  |  |  |  |  |

이때 와 의 분포렬을 구하시오.

2. 우연량 의 밀도함수가 이다. 이때 우연량 의 확률분포를 구하시오. 여기서 

3. 우연량 에 대하여 의 분포렬을 구하시오.

4. 우연량 에 대하여 의 분포를 구하시오.

5. 우연량 에 대하여 우연량 의 밀도함수 를 구하시오.

6. 원의 직경이 구간 (0,1)에서의 평등분포에 따른다고 할 때 원면적의 밀도함수를 구하시오.

7. 우연량 에 대하여 의 밀도함수를 구하시오.

8. 우연량 에 대하여

① 의 밀도함수를 구하시오.

② 를 구하시오.

9. 우연량 에 대하여

① 를 구하시오.

② 의 밀도함수를 구하시오.

10. 우연량 에 대하여 다음과 같은 의 밀도함수를 구하시오.

①  ②  ③ ④ 

11. 우연량 의 밀도함수가 다음과 같다.



이때 다음 우연량들의 분포를 구하시오.

①  ②  ③ 

12. 우연량 에 대하여 의 분포를 구하시오.

13. 우연량 에 대하여 의 수학적기대값과 분산을 구하시오.

14. 우연량 에 대하여 다음과 같은 의 밀도함수를 구하시오.

①  ② 

15. 우연량 의 밀도함수가 다음과 같다.



이때 다음과 같은 의 밀도함수를 구하시오.

①  ②  ③ 

16. 우연량 가 파라메터가 2인 지수분포에 따른다고 할 때 와 는 모두 구간 (0,1)우의 평등분포에 따른다는것을 증명하시오.

17. 우연량 에 대하여 를 증명하시오.

18. 우연량 에 대하여 를 구하시오.

## 2.7분포의 기타 특징수

수학적기대값과 분산은 우연량의 가장 중요한 특징수이다. 이밖에 우연량에는 다른 특징수들이 있는데 아래에서 그것들에 대한 정의를 내리고 그 의미를 해석한다.

### 2.7.1 차모멘트

정의 2.7.1 를 우연량, 를 정의 옹근수라고 하자. 만일 다음과 같은 수학적기대값들이 모두 존재한다면



를 의 **차원시모멘트**라고 하며



를 의 **차중심모멘트**라고 한다.

명백한것은 1차원시모멘트는 수학적기대값이고 2차중심모멘트는 분산이라는것이다. 이므로 차모멘트가 존재할 때 차모멘트가 존재하며 따라서 차이하의 모든 모멘트들이 존재한다.

중심모멘트와 원시모멘트사이에는 다음과 같은 단순한 관계가 있다.

.

4차까지의 중심모멘트들을 원시모멘트에 관하여 표시할수 있다.



실례2.7.1 우연량 에 대하여

이다.

가 홀수일 때 피적분함수는 홀함수이다. 따라서.

가 짝수일 때 피적분함수는 짝함수이며 변환 을 리용하면



그러므로 분포의 4차까지의 원점모멘트는 다음과 같다.

.

또한 이므로 원시모멘트와 중심모멘트는 같다. 즉 

### 2.7.2 변이곁수

분산이나 표준편차는 우연량이 취하는 값의 변동정도를 반영하지만 두 우연량의 변동크기를 비교할 때 분산 이나 표준편차의 크기만을 고려하면 불합리한 현상이 나타나기도 한다. 그 원인은 두가지인데 첫째로, 우연량들이 취하는 값의 단위가 다를 때 분산이나 표준편차로 그 변동의 크기를 비교하는것이 합리적이지 못하며, 둘째로, 우연량들의 단위가 같은 경우에 취하는 값의 크기는 상대적인 문제로서 큰 값을 가지는 우연량의 분산이나 표준편차는 좀 클수 있다.

그러므로 두 우연량의 변동크기를 비교할 때 어떤 경우에는 아래에 정의하는 변이계수를 리용하는것이 더 좋다.

정의 2.7.2 우연량 의 2차모멘트가 존재할 때 다음과 같은 비값을 의 **변이곁수**라고 한다.

.

표준편차와 수학적기대값의 단위가 일치하기때문에 변이곁수는 무본위량이며 따라서 단위가 변동에 미치는 영향을 제가한다.

실례 2.7.2 가 같이 자란 나무의 높이(m단위)를, 가 나이가 같은 아이들의 키(m단위)를 표시하고 이라고 하자. 이라는 사실로부터 의 변동이 작다로 볼수 있는가?

풀기. 여기서 값의 상대적크기가 문제로 되므로 변이곁수를 리용하여 비교하여야 합리적이다. 의 변이곁수는 각각 다음과 같다.

.

따라서 의 변동이 의 변동보다 크다.

### 2.7.3 분위수

정의 2.7.3 련속우연량 의 분포함수는 , 밀도함수는 이다. 임의의 에 대하여 다음의 조건을 만족하는 를 이 분포의 분위수 또는 아래쪽분위수라고 한다.

.

많은 확률통계문제는 마지막에 모두 확률부등식 를 만족시키는 최대의 를 구하는데 귀착되며 이 풀이를 를 리용하여 표시할수 있다. 그리하여 흔히 리용하는 분포(정규분포, 분포, 분포 등)에 대한 여러가지 분위수표(부록 참고)를 작성하여 리용하고있다.

그림 2.7.1(a)에서 볼수 있는것처럼 분위수 는 밀도함수아래의 면적을 두 부분으로 나누는데 왼쪽면적이 바로 이다.

마찬가지로 다음의 조건을 만족시키는 을 이 분포의 웃쪽분위수라고 한다.



그림 2.7.1(b)에서 볼수 있는것처럼 웃쪽분위수 는 밀도함수아래의 면적을 두 부분으로 나누는데 오른쪽면적이 바로 이다.

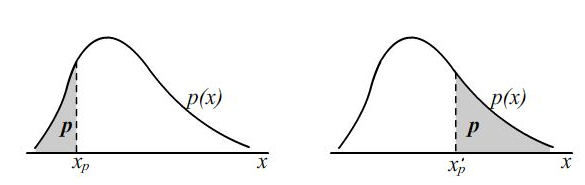


그림 2.7.1분위수와 웃쪽분위수의 차이

(b) 웃쪽분위수

(a) 아래쪽분위수

아래쪽분위수와 웃쪽분위수의 차이점을 잘 구분하여야 한다.

분위수와 웃쪽분위수는 서로 전환할수 있으며 그 공식은 다음과 같다.



실례 2.7.3 표준정규분포 의 분위수를 로 표시한다. 이것은 방정식 의 유일풀이로서 이며 여기서 는 표준정규분포함수의 거꿀함수이다. 부록 2의 표준정규분포함수표를 리용하면 로부터 를 구할수 있다. 실례로 . 일반적으로 리용되는 표준정규분포의 분위수는 그리 많지 않은데 그것을 표 2.7.1에 주었다. 표준정규분포의 밀도함수가 짝함수이므로 그 분위수는 다음과 같은 성질을 가진다.

• 이면 이다.

• 이면 이다.

• 이면 이다.

• 일 때 이다.

• 임의의 에 대하여 이다. 실례로 이다.

표 2.7.1 표준정규분포의 분위수(일부)

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 0.999 | 0.995 | 0.990 | 0.975 | 0.950 | 0.900 | 0.850 | 0.800 |
|  | 3.090 | 2.576 | 2.326 | 1.960 | 1.645 | 1.282 | 1.037 | 0.842 |

그리고 정리 2.5.1로부터 일반정규분포 의 분위수 는 방정식 의 풀이라는것을 알수 있다. 따라서 로부터 와 의 다음과 같은 관계식을 얻을수 있다.

.

례를 들어 정규분포 의 0.975분위수는 이다.

실례 2.7.4 어떤 베아링의 수명이 이고 가 그 수명분포의 분위수라면 시간은 이런 베아링들의 90%가 수명이 1000시간을 초과한다는것을 의미한다. 다른 베아링의 수명이 이고 분위수를 라고 표시하면 시간일 때 0.1의 분위수로 설명하면 후자의 질이 전자의 질보다 높다는것을 의미한다.

### 2.7.4 중위수

정의 2.7.4 련속우연량 의 분포함수는 이고 밀도함수는 라고 하자. 의 분위수 를 이 분포의 **중위수**라고 한다. 즉 는 다음과 같은 식을 만족한다.



중위수의 위치는 그림 2.7.2에서와 같이 항상 분포의 중간에 있다.

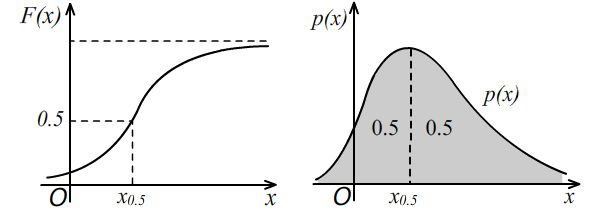


그림 2.7.2 련속우연량의 중위수

리산분포의 경우에도 분위수와 중위수의 개념을 도입할수 있지만 어떤 경우에는 주어진 에 대한 가 존재하지 않거나 유일하지 않을수 있다. 그러므로 리산분포의 경우에는 분위수가 거의 리용되지 않는데 여기서는 더 이상 론하지 않는다.

중위수는 평균과 마찬가지로 우연량의 위치에 대한 특징수이지만 중위수가 평균보다 더 잘 설명하는 경우도 있다. 례를 들면 학급학생들의 시험성적의 중간수가 80점이면 학급학생절반의 성적이 80점보다 낮고 나머지 절반의 성적은 80점이상이라는것을 의미한다. 시험성적의 평균이 80점이라면 이런 결론을 얻을수 없다. 다른 실례를 든다면 를 각각 두 나라 A, B사람들의 년령, 를 의 중위수라고 할 때 는 A나라사람들중 절반의 년령이 40살미만이고 나머지 절반은 40살이상이라는것, 은 B나라가 더 로령화되였다는것을 말하여준다.

실례 2.7.5 지수분포 의 중위수 는 방정식 의 풀이로서 이다.

가령 한 도시의 전화통화시간 (분)가 평균값 (분)인 지수분포에 따를 경우 이므로 중위수는 분이다.

이것은 도시 전체 통화의 약 절반이 1.39분안에 끝났으며 나머지 절반은 1.39분이상 걸렸다는것을 의미한다.

실례2.7.6 련속우연량 의 밀도함수가 다음과 같다.



이 분포의 0.95분위수 와 중위수 를 구하시오.

풀기. 의 분포함수가



이므로 이고 따라서 . 그러므로 . 마찬가지로 로부터 이고 을 얻는다.

### 2.7.5 비대칭도

정의 2.7.5 우연량 에 대하여 3차까지의 모멘트들이 존재한다고 하면 3차중심모멘트와 표준편차의 3제곱의 비 를 의 **비대칭도**(skewness)라고 한다. 일 때 이 분포를 **정의 비대칭** 또는 **오른쪽비대칭**이라고 하며 일 때 **부의 비대칭** 또는 **왼쪽비대칭**이라고 한다.

비대칭도 는 분포가 대칭성에서 리탈되는 정도를 설명하는 하나의 특징수인데 이것은 다음과 같은 몇가지 측면에서 알수 있다.

① 밀도함수 가 수학적기대값에 관하여 대칭일 때 즉 일 때 그의 3차중심모멘트 은 반드시 0이고 따라서 이다. 례를 들어 에 관하여 대칭인 정규분포 에 대하여 비대칭도는 모두 0이다.

② 일 때 이 분포는 비대칭분포이며 일반적으로 비대칭인 두개의 꼬리부를 가진다. 그림 2.7.3에서 볼수 있는것처럼 중꼬리(heavy tail)가 오른쪽에 있을 때 즉 중심에서 리탈되는 변수들의 값이 비교적 큰 경우 이며 오른쪽비대칭분포라고 부른다. 반대로 중꼬리가 왼쪽에 있을 때 이며 왼쪽비대칭분포라고 부른다.

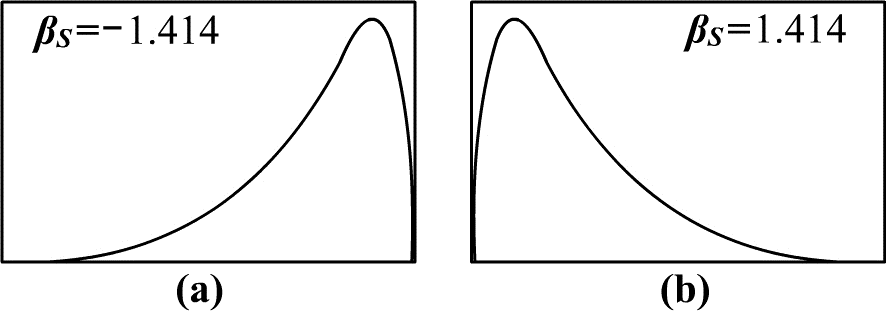


그림 2.7.3 두 밀도함수, (a): 왼쪽비대칭, (b): 오른쪽비대칭.

③ 비대칭도 는 표준편차의 3제곱 을 단위로 하여 3차중심모멘트의 크기를 측정하므로 무본위량이고 비교가 합리적이다. 간단히 말하면 분포의 3차중심모멘트 가 비대칭도의 부호를 결정하고 분포의 표준편차 가 그 크기를 결정한다.

실례 2.7.7 3개의 베타분포 의 비대칭도를 구하시오.

풀기. 우연량 가 베타분포 에 따른다고 하면 3차까지의 원점모멘트를 계산할수 있다.



의 계산상 편리를 위하여 를 웃식들에 대입하면 3차까지의 중심모멘트들을 계산할수 있다.

.

그러면 베타분포의 비대칭도는 이다. 을 대입하면

의 이고 오른쪽비대칭이다.

의 이고 대칭이다.

의 이고 왼쪽비대칭이다.

### 2.7.6 뾰족도

정의 2.7.6 우연량 의 4차까지의 모멘트가 모두 존재한다면 를 의 **뾰족도**(kurtosis)라고 부른다.

뾰족도는 분포의 뾰족한 정도와 꼬리의 굵기를 설명하는 특징수로서 다음과 같은 몇가지 측면에서 알아볼수 있다.

① 정규분포 에서 이고 정의에 따르면 모든 정규분포에서 이다. 그러므로 여기서 론의하는 뾰족도는 일반 밀도함수의 봉우리가 높고 낮음을 나타내지 않는다. 그것은 정규분포 의 봉우리값은 로서 표준편차 와 반비례하며 가 작을수록 정규분포의 봉우리값이 크기때문이다. 그런데 이 경우 뾰족도는 와는 관계가 없다.

② 우의 정의에서 분자와 분모를 각각 로 나누고 의 표준화된 변수를 로 표시하면 를 로 쓸수 있다. 여기서 이고 는 표준정규분포우연량이며 이다.

이로부터 뾰족도 는 정규분포에 비한 초과량이며 즉 의 표준화된 변수와 표준정규우연량의 4차원점모멘트들사이의 차이며 표준정규분포를 기준으로 그 크기를 결정한다.

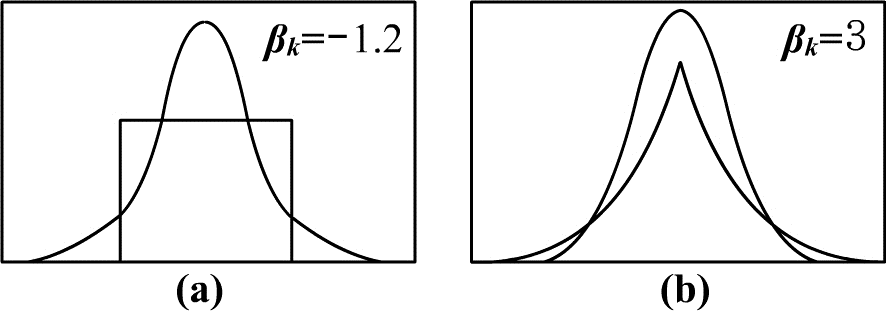
• 은 표준화된 분포가 표준정규분포보다 뾰족하고 꼬리가 더 굵다는것을 의미한다(그림 2.7.4(b)에서 뾰족한 곡선은 라플라스분포이다).

• 은 표준화된 분포가 표준정규분포보다 더 평탄하고 꼬리가 더 가늘어진다는것을 의미한다 (그림 2.7.4(a)에서 직4각형모양의 곡선은 평등분포이다)

• 은 표준화된 분포의 뾰족정도와 꼬리의 굵기정도가 표준정규분포와 같다는것을 의미한다.

그림 2.7.4 두개의 밀도함수와 표준정규분포밀도함수의 비교

이것들은 평균과 분산이 같고 비대칭도가 0(대칭분포)이지만 뾰족도는 크게 차이난다.



③ 비대칭도와 뾰족도는 모두 분포형태를 서술하는 특징수이다. 이것들은 모두 정규분포를 기준으로 하며 정규분포의 비대칭도와 뾰족도는 모두 0이다. 실천에서 어떤 분포의 비대칭도와 뾰족도가 모두 0일 때 또는 거의 0일 때 그 분포는 정구분포 또는 근사정규분포로 본다.

④ 여러 일반적인 분포의 비대칭도와 뾰족도를 표 2.7.2에 주었다. 감마분포 의 비대칭도와 뾰족도는 와만 관련되고 와는 관련이 없으므로 를 형태파라메터라고 하지만 는 형태파라메터라고 하지 않는다. 평등분포 와 지수분포 의 비대칭도와 뾰족도는 모두 파라메터들과 관계가 없으므로 들은 모두 형태파라메터라고 할수 없다. 베타분포 의 비대칭도와 뾰족도는 그 파라메터 와 관련되며 따라서 이것들을 형태파라메터라고 부를수 있다.

표 2.7.2여러 일반적인 분포의 비대칭도와 뾰족도

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 분포 | 평균값 | 분산 | 비대칭도 | 뾰족도 |
| 평등분포 |  |  | 0 | -1.2 |
| 정규분포 |  |  | 0 | 0 |
| 지수분포 |  |  | 2 | 6 |
| 감마분포 |  |  |  |  |

실례2.7.8 감마분포 의 비대칭도와 뾰족도를 계산하시오.

풀기. 감마분포 의 차 원점모멘트를 계산한다.

.

4차까지의 원점모멘트는 일 때 얻을수 있다.



이것을 리용하여 2차, 3차, 4차 중심모멘트를 얻는다.



마지막으로 의 비대칭도와 뾰족도를 얻을수 있다.



따라서 감마분포의 비대칭도는 에 반비례하고 뾰족도는 에 반비례한다. 가 클수록 와 를 0에 가깝게 할수 있고 이때 감마분포는 점점 더 정규분포에 가깝다.

련습문제 2.7

1. 우연량 가 있을 때 에 대하여 를 구하고 이 분포의 비대칭도와 뾰족도를 구하시오.

2. 우연량 에 대하여 변이곁수를 구하시오.

3. 다음과 같은 분포의 중위수를 구하시오.

① 구간 에서의 평등분포 ② 정규분포  ③ 로그정규분포 

4. 우연량 가 있을 때 에 대하여 를 구하시오.

5. 우연량 가 있을 때 에 대하여 를 구하고 이 분포의 변이곁수, 비대칭도, 뾰족도를 구하시오.

6. 우연량 가 정규분포 에 따른다고 할 때 과 를 구하시오.

7. 우연량 가 두파라메터웨이블분포에 따르고 그 분포함수가 다음과 같다고 하자.



여기서 이다. 이 분포함수의 분위수 의 표시식을 구하고 일 때의 을 구하시오.

8. 자유도가 2인 분포의 밀도함수가 다음과 같다.



이때 분포함수와 분위수 을 구하시오.

9. 우연량 의 분포밀도함수 가 직선 에 관하여 대칭이고 가 존재한다고 가정하자. 이때

① 대칭점 는 평균값이면서 중위수 즉 라는것을 증명하시오.

② 이면 라는것을 증명하시오.

10. 우연량 의 비대칭도와 뾰족도는 평행이동과 척도변환에 대하여 불변 즉 임의의 실수 에 대하여 와 는 같은 비대칭도와 뾰족도를 가진다는것을 증명하시오.

11. 어떤 기계의 보수시간 (분)는 로그정규분포 에 따른다. 이때

① 분위수 를 구하시오.

② 일 때 이 분포의 중위수를 구하시오.

③ 일 때 보수를 95%완성하는 시간을 구하시오.

12. 어떤 절연재료의 수명 (시간)는 로그정규분포 에 따른다. 분위수 , 시간이 주어졌을때 와 를 구하시오.

13. 어떤 공장에서 과거의 생산실적에 따라 월생산액이 가장 높은 5%의 로동자들에게 상금을 주기로 결정하였다. 일인당 월생산액 (kg)이 정규분포 에 따른다면 상금수여기준을 생산액의 얼마로 정해야 하겠는가?