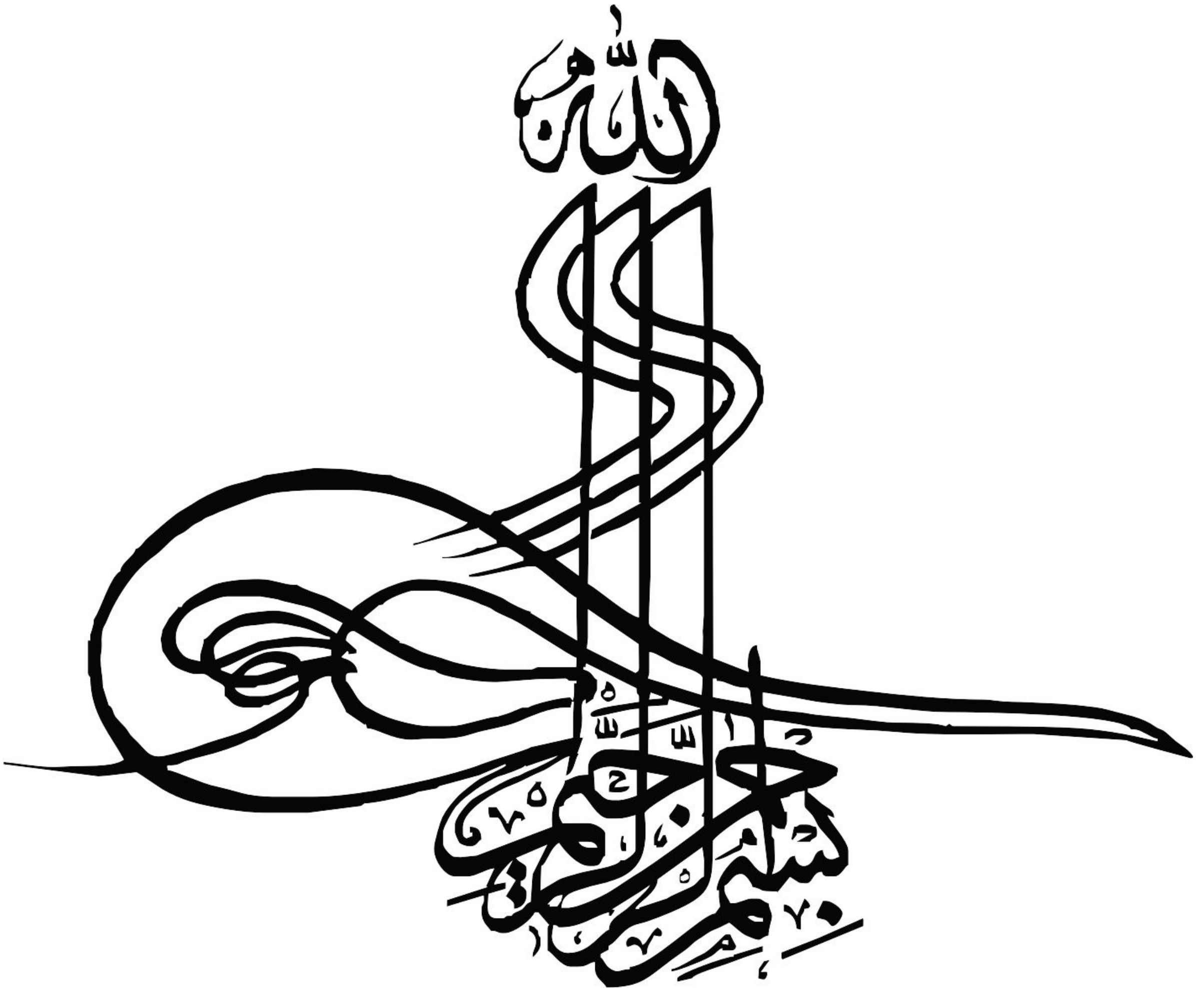
بسم الله الرحمن الرحيم

دانشکده علوم ریاضی
گروه ریاضی
(گرایش کاربردی)

عنوان :

مولد اعداد شبه‌تصادفی فائور و کاربرد آن

از:

الهام رادمقدم

استاد راهنما:

دکتر بهروز فتحی واجارگاه

شهریور ۱۳۹۱

تقدیم به

پدر و مادر عزیزم

خدای را بسی شاکرم که از روی کرم، پدر و مادری فداکار نصیبم ساخته تا در سایه درخت پربار وجودشان بیاسایم و از ریشه آنها شاخ و برگ گیرم و از سایه وجودشان در راه کسب علم و دانش تلاش نمایم.

والدینی که بودنشان تاج افتخاری است بر سرم و نامشان دلیلی است بر بودنم چراکه این دو وجود پس از پروردگار، مایه هستی ام بوده اند دستم را گرفتند و راه رفتن را در این وادی زندگی پر از فراز و نشیب آموختند.

آموزگارانی که برایم زندگی، بودن و انسان بودن را معنا کردند.

حال این برگ سبزی است تحفه درویش تقدیم آنان....

پروردگارا

ای هستی بخش وجود، مرا با نعمات بی کرانت ، توان شکر نیست . الهی مرا مدد کن تا دانش اندکم ، نه نردبانی برای فزونی تکبر و غرور و نه حلقه ای برای اسارت و نه دست مایه ای برای تجارت، بلکه تکیه گاهی باشد برای تجلیل از تو.

تشکر می کنم از آنان که در راه اعتلای علمی ایران از هیچ کوششی دریغ نمی ورزند.

مراتب تقدیر و قدردانی خود را از استاد ارجمند، جناب آقای دکتر بهروز فتحی اعلام می دارم.

به خاطر زحمات بی دریغتان در امر راهنمائی و تدریس توام با اخلاق و رفتاری نیکو، از شما کمال تشکر را دارم و بر این اعتقادم که همواره اندیشه های علمی و نظرات محققانه شما راه گشای آینده ی ما خواهد بود.

مراتب قدردانی خود را از نظرات ارزشمند اساتید محترم ، داوران این پایان نامه ، جناب آقای دکتر امیر زینل و جناب آقای دکتر فرشید مهردوست و همچنین از مدیر محترم گروه ریاضی جناب آقای دکتر کیان پور ، اعلام می دارم.

همچنین از تمامی دوستانی که بی منت مرا یاری رساندند، صمیمانه سپاسگزارم.

فهرست مطالب

لیست جداول ج

لیست تصاویر چ

چکیده فارسی خ

چکیده انگلیسی د

پیشگفتار ۱

۱ مفاهیم پایه‌ای احتمال و فرایند ۲

۱-۱ احتمال ۳

۱-۲ متغیرهای تصادفی، توابع توزیع و امیدریاضی ۴

۱-۳ خانواده‌های پارامتری ویژه از توزیع‌های یک متغیره ۶

۱-۳-۱ توزیع‌های گسسته ۷

۱-۳-۲ توزیع‌های پیوسته ۸

۱-۴ کوواریانس و ضریب همبستگی، احتمال شرطی و استقلال پیشامدها ۱۰

۱-۴-۱ کوواریانس و ضریب همبستگی ۱۰

۱-۴-۲ احتمال شرطی و استقلال پیشامدها ۱۰

۱-۵ نمونه‌گیری و توزیع‌های نمونه‌گیری ۱۳

۱-۶ فرایند تصادفی ۱۷

۱-۶-۱ قدم‌زدن تصادفی ۱۸

۱-۶-۲ زنجیرهای مارکف ۱۸

۲ شبیه‌سازی ۲۰

۲-۱ اعداد تصادفی ۲۱

۲-۱-۱ تولید اعداد تصادفی‌نما ۲۱

۲-۱-۲ محاسبه‌ی انتگرال‌ها با استفاده از اعداد تصادفی ۲۲

۲-۱-۳ برآورد π ۲۳

۲-۲ تولید متغیرهای تصادفی گسسته ۲۶

۲-۲-۱ روش تبدیل وارون ۲۶
۲-۳ تولید متغیرهای تصادفی پیوسته ۲۹
۲-۳-۱ تولید متغیر تصادفی پواسن ۲۹
۲-۳-۲ تکنیک پذیرش-عدم پذیرش ۲۹

۳ دنباله‌ی فائور و نتایج عددی **۳۱**
۳-۱ انتگرال مونت‌کارلو و شبه مونت‌کارلو ۳۲
۳-۱-۱ انتگرال مونت‌کارلو ۳۲
۳-۱-۲ مکانیزم انتگرال مونت‌کارلوی استاندارد ۳۳
۳-۱-۳ انتگرال شبه مونت‌کارلو ۳۴
۳-۱-۴ دنباله‌های با اختلاف پایین (شبه تصادفی (QMC)) ۳۵
۳-۲ بعد موثر و تعیین آن ۳۹

۴ اسکرمبل‌های دنباله‌ی فائور و آزمون انتگرال **۴۷**
۴-۱ روش‌های شبه تصادفی اسکرمبل‌شده ($RQMC$) ۴۸
۴-۱-۱ دنباله‌ی جی‌فائور ($GFaure$) ۴۸
۴-۱-۲ اسکرمبل (رقمی) خطی تصادفی ۴۸
۴-۱-۳ اسکرمبل دوجمله‌ای-I ($I - binomial$) ۴۹
۴-۱-۴ اسکرمبل استریپد ($striped$) ۵۰
۴-۱-۵ اسکرمبل غیرخطی ۵۰
۴-۱-۶ اسکرمبل جدید ۵۱
۴-۲ انتگرال‌های عددی ۵۲

آ برنامه‌های کامپیوتری **۶۲**

ب جدول ها **۸۵**

منابع و مآخذ **۹۱**

واژه‌نامه انگلیسی به فارسی **۹۴**

لیست جداول

ب-۱ خطای نسبی انتگرال $1 = \int_0^1 \cdots \int_0^1 \prod_{i=1}^{s} \frac{\pi}{2}|4x_i - 2|dx_1...dx_s$ نسبت به تعداد اعضای نمونه در بعد۴. . . . . ۸۶

ب-۲ خطای نسبی انتگرال $1 = \int_0^1 \cdots \int_0^1 \prod_{i=1}^{s} \frac{\pi}{2}|4x_i - 2|dx_1...dx_s$ نسبت به تعداد اعضای نمونه در بعد۷. . . . . ۸۶

ب-۳ خطای نسبی انتگرال $1 = \int_0^1 \cdots \int_0^1 \prod_{i=1}^{s} \frac{\pi}{2}|4x_i - 2|dx_1...dx_s$ نسبت به تعداد اعضای نمونه در بعد۱۳. . . . . ۸۶

ب-۴ خطای نسبی انتگرال $1 = \int_0^1 \cdots \int_0^1 \prod_{i=1}^{s} \frac{\pi}{2}|4x_i - 2|dx_1...dx_s$ نسبت به تعداد اعضای نمونه در بعد ۲۰. . . . . ۸۷

ب-۵ خطای نسبی انتگرال $1 = \int_0^1 \cdots \int_0^1 \prod_{i=1}^{s} \frac{\pi}{2}|4x_i - 2|dx_1...dx_s$ نسبت به تعداد اعضای نمونه در بعد ۲۵. . . . . ۸۷

ب-۶ خطای نسبی انتگرال $1 = \int_0^1 \cdots \int_0^1 \prod_{i=1}^{s} \frac{\pi}{2}|4x_i - 2|dx_1...dx_s$ نسبت به تعداد اعضای نمونه در بعد ۴۰. . . . . ۸۷

ب-۷ زمان اجرای برنامه‌ی خطای انتگرال $1 = \int_0^1 \cdots \int_0^1 \prod_{i=1}^{s} \frac{\pi}{2}|4x_i - 2|dx_1...dx_s$ نسبت به تعداد اعضای نمونه در بعد۴. . . . . ۸۸

ب-۸ زمان اجرای برنامه‌ی خطای انتگرال $1 = \int_0^1 \cdots \int_0^1 \prod_{i=1}^{s} \frac{\pi}{2}|4x_i - 2|dx_1...dx_s$ نسبت به تعداد اعضای نمونه در بعد۷. . . . . ۸۸

ب-۹ زمان اجرای برنامه‌ی خطای انتگرال $1 = \int_0^1 \cdots \int_0^1 \prod_{i=1}^{s} \frac{\pi}{2}|4x_i - 2|dx_1...dx_s$ نسبت به تعداد اعضای نمونه در بعد ۱۳. . . . . ۸۸

ب-۱۰ زمان اجرای برنامه‌ی خطای انتگرال $1 = \int_0^1 \cdots \int_0^1 \prod_{i=1}^{s} \frac{\pi}{2}|4x_i - 2|dx_1...dx_s$ نسبت به تعداد اعضای نمونه در بعد ۲۰. . . . . ۸۹

ب-۱۱ زمان اجرای برنامه‌ی خطای انتگرال $1 = \int_0^1 \cdots \int_0^1 \prod_{i=1}^{s} \frac{\pi}{2}|4x_i - 2|dx_1...dx_s$ نسبت به تعداد اعضای نمونه در بعد ۲۵. . . . . ۸۹

ب-۱۲ زمان اجرای برنامه‌ی خطای انتگرال $1 = \int_0^1 \cdots \int_0^1 \prod_{i=1}^{s} \frac{\pi}{2}|4x_i - 2|dx_1...dx_s$ نسبت به تعداد اعضای نمونه در بعد ۴۰. . . . . ۸۹

ب-۱۳ خطای انتگرال $1 = \int_0^1 \cdots \int_0^1 \prod_{i=1}^{s} \frac{\pi}{2} sin(\pi x_i)dx_1...dx_s$ نسبت به تعداد اعضای نمونه. ۹۰

ب-۱۴ زمان اجرای برنامه‌ی خطای انتگرال $1 = \int_0^1 \cdots \int_0^1 \prod_{i=1}^{s} \frac{\pi}{2} sin(\pi x_i)dx_1...dx_s$ نسبت به تعداد اعضای نمونه. ۹۰

لیست تصاویر

۲-۱ . . . . ۲۳
۲-۲ . . . . ۲۳
۲-۳ هیستوگرام تابع رند متلب برای N=۱۰۰۰. ۲۴
۲-۴ هیستوگرام تابع پی رند متلب برای N=۱۰۰۰. . . . . ۲۵
۲-۵ خطای مطلق انتگرال $I(f) = \int_۰^۱ \sqrt{۱ - x^۲}dx$ نسبت به تعداد اعضای نمونه، برای دو تابع رند و پی رند ۲۶

۳-۱ . . . . ۳۲
۳-۲ خطای مطلق انتگرال $I(f) = \int_۰^۱ \sqrt{۱ - x^۲}dx$ نسبت به تعداد اعضای نمونه، برای تابع رند و دنباله‌ی فائور ۳۸
۳-۳ خطای مطلق انتگرال $I(f) = \int_۰^۱ \frac{\pi}{۲} sin(\frac{\pi}{۲}(x + \frac{\pi}{۴}))dx$ نسبت به تعداد اعضای نمونه، برای تابع رند و دنباله‌ی فائور ۳۹
۳-۴ دنباله‌ی هالتون بعد۱و۲ (N=۱۰). ۴۳
۳-۵ دنباله‌ی هالتون بعد ۱و۲ (N=۱۰۰۰). ۴۳
۳-۶ دنباله‌ی هالتون بعد ۱۴و۱۵ (N=۱۰۰۰). ۴۴
۳-۷ دنباله‌ی هالتون بعد ۲۷و۲۸ (N=۱۰۰۰). ۴۴
۳-۸ دنباله‌ی فائور در پایه‌ی۳ و بعد ۱و۲ (N=۱۰۰۰). ۴۵
۳-۹ دنباله‌ی فائور در پایه‌ی۱۷ و بعد ۱۳و۱۴ (N=۱۰۰۰). ۴۵
۳-۱۰ دنباله‌ی فائور در پایه‌ی۳۱ و بعد ۲۸و۲۹ .(N=۱۰۰۰) ۴۶
۳-۱۱ ۱۰۲۴ نقطه از دنباله‌ی فائور با بعد۴۰ و در پایه‌ی ۴۱. . . . . ۴۶

۴-۱ ۱۰۲۴ نقطه از دنباله‌ی جی‌فائور با بعد۴۰ و در پایه‌ی ۴۱. . . . . ۵۵
۴-۲ ۱۰۲۴ نقطه از دنباله‌ی فائور با اسکرمبل خطی رقمی با بعد۴۰ و در پایه‌ی ۴۱. . . . . ۵۵
۴-۳ ۱۰۲۴ نقطه از دنباله‌ی فائور با اسکرمبل خطی با بعد۴۰ و در پایه‌ی ۴۱. . . . . ۵۶
۴-۴ ۱۰۲۴ نقطه از دنباله‌ی فائور با اسکرمبل دوجمله‌ای با بعد۴۰ و در پایه‌ی ۴۱. . . . . . . ۵۶
۴-۵ ۱۰۲۴ نقطه از دنباله‌ی فائور با اسکرمبل استریپد با بعد۴۰ و در پایه‌ی ۴۱. . . . . ۵۷
۴-۶ ۱۰۲۴ نقطه از دنباله‌ی فائور با اسکرمبل پیش‌معکوس با بعد۴۰ و در پایه‌ی ۴۱. . . . . . . ۵۷
۴-۷ ۱۰۲۴ نقطه از دنباله‌ی ام‌فائور با بعد۴۰ و در پایه‌ی ۴۱. . . . . ۵۸
۴-۸ ۱۰۲۴ نقطه از دنباله‌ی ام۲ فائور با بعد۴۰ و در پایه‌ی ۴۱. . . . . ۵۸
۴-۹ خطای نسبی نسبت به تعداد اعضای نمونه برای آزمون انتگرال (۱). ۵۹

۴-۱۰ خطای نسبی نسبت به تعداد اعضای نمونه برای آزمون انتگرال (۲) ۵۹
۴-۱۱ خطای نسبی نسبت به تعداد اعضای نمونه برای آزمون انتگرال (۳) ۶۰
۴-۱۲ نمودار زمان اجرای برنامه‌ی خطای آزمون انتگرال (۱) نسبت به تعداد اعضای نمونه. ۶۰
۴-۱۳ نمودار زمان اجرای برنامه‌ی خطای آزمون انتگرال (۲) نسبت به تعداد اعضای نمونه. ۶۱
۴-۱۴ نمودار زمان اجرای برنامه‌ی خطای آزمون انتگرال (۳) نسبت به تعداد اعضای نمونه. ۶۱

چکیده:

مولد اعداد شبه تصادفی فائور و کاربرد آن

الهام رادمقدم

هدف ما در این پایان‌نامه، معرفی دنباله‌ی شبه‌تصادفی فائور و دنباله‌های اسکرمبل‌شده‌ی آن می‌باشد. در پایان یک الگوریتم جدید ارائه می‌دهیم و کارایی آن را با دنباله‌های اسکرمبل‌شده‌ی قبلی فائور در مسائل انتگرالی، مقایسه می‌کنیم.

کلید واژه:

روش شبه‌مونت‌کارلو، مولدهای شبه‌تصادفی، دنباله‌ی فائور، اسکرمبل کردن دنباله‌ی فائور، انتگرال عددی

Abstract:

Faure quasi random numbers generator and its application

Elham Radmoghaddam

In this dissertation, we introduce the Faure sequence and scrambled the Faure quasi-random number sequence. Finally, we present a new algorithm and compare its efficiency with previous scrambled Faure quasi-random number sequence for integration problems.

Key words:

Quasi Monte Carlo method, Quasi-random number generator, Faure sequence, Scrambling Faure sequence, Numerical integration

پیشگفتار

انتگرال‌های با بعد بالا کاربردهای فراوانی دارند. روش‌های مونت‌کارلو با موفقیت برای انتگرال‌های با بعد بالا، استفاده می‌شوند. برای دست یافتن به جواب چنین انتگرال‌هایی، باید به‌دنبال الگوریتمی باشیم که هم مقدار عددی انتگرال را تخمین بزند و هم قادر به برآورد مقدار خطا باشد. این الگوریتم باید در حد امکان با کم‌ترین هزینه‌ی محاسباتی جواب قابل قبولی را ارائه دهد. اساس روش مونت‌کارلو نمونه‌گیری تصادفی است که بر پایه‌ی تولید اعداد تصادفی صورت می‌گیرد. انتگرال چندگانه به بعد وابسته است، ساخت مربع کلاسیک برای ابعاد بالاتر از ۵ یا ۶ امکان‌پذیر نیست. به این معنا که تجسم مربع واحد برای ابعاد بالاتر از ۵ امکان‌پذیر نیست. روش‌های مونت‌کارلوی استاندارد ابزاری مفید برای حل این مشکل هستند. این روش‌ها از میانگین نمونه برای تقریب زدن انتگرال استفاده می‌کنند. روش‌های شبه‌تصادفی نسخه‌ی تعیین‌شده‌ی روش‌های استاندارد هستند. دنباله‌های شبه‌تصادفی دارای اختلاف پایین‌تری نسبت به روش استاندارد هستند[۳۱].

اعداد شبه‌تصادفی به‌عنوان تعمیمی از اعداد تصادفی، در الگوریتم‌ها به‌کار می‌روند.

در این پایان‌نامه، ابتدا به معرفی مختصری از روش‌های مونت‌کارلو و شبه‌مونت‌کارلو می‌پردازیم. مزایا و معایب هر روش را بیان می‌کنیم. سپس ضمن بررسی مولدهای شبه‌تصادفی نظیر دنباله‌ی فائور، از این دنباله برای حل انتگرال‌های با بعد بالا استفاده می‌کنیم. الگوریتم‌ها توسط نرم افزار $Matlab$ پیاده‌سازی شده‌اند.

فصل ۱

مفاهیم پایه‌ای احتمال و فرایند

مقدمه

در این فصل، ما به‌طور مختصر به مفاهیم پایه‌ای احتمال و فرایندهای تصادفی می‌پردازیم. ما از این مفاهیم برای مدل‌سازی مسائل مورد نظر در روش مونت‌کارلو استفاده خواهیم کرد.

۱-۱ احتمال

یکی از اهداف علم، پیش‌بینی و توصیف پیشامدها در دنیایی است که در آن زندگی می‌کنیم. یک راه برای رسیدن به این منظور، ساختن مدل‌های ریاضی است که دنیای واقعی را به‌طور مناسبی توصیف کند. به عنوان مثال معادله‌ی $s = \frac{1}{2}gt^2$ رابطه‌ی معینی بین نمادهای t, g, s را بیان می‌کند. این معادله‌یک مدل ریاضی است. در استفاده از معادله‌ی $s = \frac{1}{2}gt^2$ که تابعی از زمان است، برای پیش‌بینی s، مقدار مسافتی که جسم درحال سقوط می‌پیماید، باید ثابت گرانشg معلوم باشد. ثابت اخیر یک ثابت فیزیکی است که برای مفید واقع شدن در معادله‌ی $s = \frac{1}{2}gt^2$ باید به‌وسیله‌ی آزمایش، اندازه‌گیری شود. فایده‌ی ارائه‌ی این معادله این است که مدل احتمالی بسازیم که بتوان ن را برای توصیف پیشامدها در جهان واقعی مورد استفاده قرار داد.

مفهوم احتمال کلاسیک وقتی به‌کار می‌رود که نظیر آنچه در اکثر بازی‌های شانسی پیش می‌آید، تمام برآمدهای ممکن هم‌شانس باشند . عیب عمده‌ی مفهوم احتمال کلاسیک در محدود بودن قابلیت کاربردی آن است، زیرا وضعیت‌هایی موجودند که برای آنها امکاناتی را که رخ می‌دهند نمی‌توان هم‌شانس دانست. برای مثال برای پاسخ به این سوال که آیا در روز معینی باران خواهد آمد یا نه؟ با چنین وضعیتی سروکار داریم. از بین مفاهیم مختلف احتمال، پرطرفدارتر از همه مفهوم تعبیر فراوانی است که طبق آن احتمال وقوع یک پیشامد برابر نسبت دفعاتی است که پیشامدهایی از این نوع در تکرار زیاد رخ داده‌اند. دیدگاه دیگری که این‌روزها طرفداران بیشتری دارد، تعبیر احتمال‌ها به عنوان ارزیابی‌های ذهنی یا شخصی است و وقتی به‌کار می‌روند که شواهد کمی وجود دارد. روشی که در این بخش به‌کار خواهیم برد، روش اصل موضوعی است که در آن احتمال‌ها به عنوان "موجودات ریاضی" تعریف شده‌اند که مطابق با برخی قواعد خوش تعریف، رفتار می‌کنند. بنابراین هریک از مفاهیم ذکرشده، مادامی‌که این قواعد برقرارند، می‌توانند در کاربردها مورد استفاده قرار گیرند.

تعریف ۱-۱-۱. فضای نمونه

در آمار متداول است که به هر فرایند مشاهده‌یا اندازه‌گیری، عنوان آزمایش را اطلاق می‌کنند. نتایجی که از یک آزمایش به‌دست می‌آید برآمد آزمایش نامیده می‌شود. آزمایش تصادفی، یعنی آزمایشی که برآمد آن از قبل معلوم نیست. مجموعه‌ی تمام رویدادهای ممکن آزمایش را، فضای نمونه‌ی [1] آن آزمایش می‌گوییم.

[1] Sample state

فضای نمونه را با Ω نشان می‌دهیم. به عنوان مثال، فضای نمونه‌ی آزمایش پرتاب یک سکه $\Omega = \{H, T\}$ و فضای نمونه‌ای آزمایش پرتاب یک سکه تا وقوع اولین شیر $\Omega = \{H, TH, TTH, ...\}$ است.

تعریف ۱-۱-۲. پیشامد و فضای پیشامد

پیشامد زیرمجموعه‌ای از فضای نمونه است. رده‌ی تمام پیشامدهای مربوط به یک آزمایش مفروض را فضای پیشامد گوییم. یک پیشامد همواره زیرمجموعه‌ای از فضای نمونه است، اما برای فضاهای نمونه‌ای به اندازه‌ی کافی بزرگ، همه‌ی زیرمجموعه‌ها، پیشامد نخواهند بود. اگر فضای نمونه شمارا و متناهی باشد آنگاه می‌توان گفت که هر زیرمجموعه از فضای نمونه‌ای یک پیشامد است.

۱-۲ متغیرهای تصادفی، توابع توزیع و امیدریاضی

تعریف ۱-۲-۱. متغیر تصادفی

متغیر تصادفی تابعی است از فضای نمونه‌ی Ω به $\Re$، یعنی برای هر $\omega \in \Omega$

$$X(\omega) = x \in \Re$$

با تعریف متغیر تصادفی، احتمال از پیشامدها به $\Re$ منتقل می‌گردد. این تعریف امکان بررسی و مزایای فراوانی را به‌دلیل خواص فیزیکی که در $\Re$ با توجه به دو عمل جمع و ضرب وجود دارد، برقرار می‌کند و طیف وسیعی از پیشامدها را مطرح می‌سازد. متغیرهای تصادفی به دو نوع گسسته و پیوسته تقسیم می‌شوند. متغیر تصادفی X گسسته است هرگاه برد آن شمارا باشد در غیراین‌صورت متغیر تصادفی را پیوسته نامیم.

تعریف ۱-۲-۲. تابع توزیع تجمعی

تابع توزیع تجمعی متغیر تصادفی X را، که با $F_X(.)$ نشان می‌دهیم، تابعی با دامنه‌ی اعداد حقیقی و برد $[0, 1]$، که به ازای هرعدد حقیقی x در شرط $F_X(x) = P(X \leq x)$ صدق کند، تعریف می‌کنیم.

ویژگی‌های تابع توزیع تجمعی

۱.

$$F_X(-\infty) = \lim_{x \to -\infty} F_X(x) = 0 \quad , \quad F_X(+\infty) = \lim_{x \to \infty} F_X(x) = 1.$$

۲. $F_X(.)$ تابعی یکنوا و غیر نزولی است، یعنی برای $a < b$، $F_X(a) \leq F_X(b)$.

۳. $F_X(.)$ از راست پیوسته است، یعنی

$$\lim_{h \to 0} F_X(x + h) = F_X(x).$$

تعریف ۱-۲-۳. تابع چگالی گسسته

اگر X یک متغیر تصادفی گسسته با مقادیر مجزای $x_1, ..., x_n, ...$ باشد، تابعی که با $f_X(.)$ نشان داده می‌شود و به‌صورت زیر تعریف می‌شود، تابع چگالی گسسته‌ی X نامیده می‌شود.

$$f_X(x) = p(X = x_j) \qquad j = 1, 2, ...$$

مقادیر یک متغیر تصادفی گسسته اغلب نقاط جرم نامیده می‌شوند و $f_X(x_j)$ جرم مربوط به نقطه‌ی جرم x_j را نشان می‌دهد. تابع جرم احتمال، تابع فراوانی گسسته و تابع احتمال، اصطلاح‌های دیگری هستند که به‌جای تابع چگالی گسسته به‌کار می‌روند. همچنین گاهی برای توابع چگالی گسسته، به‌جای $f_X(.)$ از $P_X(.)$ استفاده می‌شود. در واقع هر تابع $f_X(.)$ دامنه‌ی اعداد حقیقی و برد $[0, 1]$ یک تابع چگالی گسسته است هرگاه

۱. برای هر $j = 1, 2, ...$ $\quad f(x_j) > 0$.

۲. برای $X \neq x_j$, $\quad j = 1, 2, ... \quad f(x) = 0$.

۳. $\sum_{j=1}^{\infty} f(x_j) = 1$.

این تعریف به ما این امکان را می‌دهد که از توابع چگالی گسسته، بدون مراجعه به متغیر تصادفی معینی بحث کنیم.

تعریف ۱-۲-۴. احتمال یک متغیر تصادفی پیوسته

اگر X یک متغیر تصادفی پیوسته باشد، تابع $f_X(.)$ در $F_X(x) = \int_{-\infty}^{x} f_X(u)du$، تابع چگالی احتمال X نامیده می‌شود. نام‌های دیگری که به‌جای تابع چگالی احتمال به‌کار می‌رود شامل نام‌های تابع چگالی، تابع چگالی پیوسته و تابع چگالی انتگرال می‌باشد. در واقع هر تابع $f(.)$ با دامنه‌ی اعداد حقیقی و برد $[0, \infty)$، یک تابع چگالی احتمال است هرگاه

۱. برای تمام Xها، $f(x) \geq 0$.

۲. $\int_{-\infty}^{\infty} f(x)dx = 1$.

تعریف ۱-۲-۵. امید ریاضی

اگرX متغیر تصادفی باشد، میانگین X را با $E(X)$ یا μ_X نشان می‌دهیم. در واقع میانگین متغیر تصادفی X، اندازه‌ای از مکان مرکزی چگالی X است. اگر X گسسته با نقاط جرم $x_1, x_2, ...$ باشد، امید ریاضی متغیر تصادفی X به‌صورت زیر است

$$E(X) = \sum_{-\infty}^{\infty} x_j f_X(x_j)$$

و اگر X پیوسته با تابع چگالی احتمال f_X باشد، امید ریاضی آن به‌صورت زیر تعریف می‌شود

$$E(X) = \int_{-\infty}^{\infty} x f_X(x) dx.$$

تعریف ۱-۲-۶. واریانس

واریانس متغیر تصادفی X، کمیتی برای اندازه‌گیری میزان پراکندگی چگالی X است و با σ_X^2 یا $Var(X)$ نشان داده می‌شود. اگر X گسسته با نقاط جرم $x_1, x_2, ...$ باشد، واریانس متغیر تصادفی X به‌صورت زیر تعریف می‌شود

$$Var(X) = \sum_j (x_j - \mu_x)^2 f_X(x_j)$$

و اگر X، پیوسته با تابع چگالی احتمال $f_X(x)$ باشد واریانس آن به‌صورت زیر تعریف می‌شود

$$Var(x) = \int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(x) dx$$

با اندکی محاسبه فرمول دیگری برای واریانس X برابر است با

$$Var(X) = E(X^2) - (E(X))^2.$$

۱-۳ خانواده‌های پارامتری ویژه از توزیع‌های یک متغیره

در این بخش چند خانواده‌ی پارامتری از چگالی‌های گسسته‌ی یک متغیره را فهرست می‌کنیم.

۱-۳-۱ توزیع‌های گسسته

تعریف ۱-۳-۱. توزیع یکنواخت گسسته

هر عضو خانواده‌ی توابع چگالی گسسته‌ی

$$f_X(x) = f(x;N) = \frac{1}{N} \qquad x = 1, 2, ..., N \qquad (1)$$

که در آن پارامتر N اعداد صحیح مثبت را اختیار می‌کند، دارای توزیع یکنواخت گسسته است. متغیر تصادفی X که دارای تابع چگالی رابطه‌ی (۱) باشد، متغیر تصادفی یکنواخت گسسته نامیده می‌شود.

اگر X دارای توزیع یکنواخت گسسته باشد، آنگاه $E(X) = \frac{N+1}{2}$ و $Var(X) = \frac{N^2-1}{12}$.

تعریف ۱-۳-۲. توزیع برنولی

گوییم متغیر تصادفی X دارای توزیع برنولی است اگر تابع چگالی گسسته‌ی X به‌صورت زیر باشد.

$$f_X(x) = f_X(x;p) = p^x(1-p)^{1-x} \qquad x = 0, 1$$

که در آن پارامتر p در $0 \leq p \leq 1$ صدق می‌کند. $1-p$ اغلب با q نشان داده می‌شود.

تعریف ۱-۳-۳. توزیع دوجمله‌ای

گوییم متغیر تصادفی X دارای توزیع دوجمله‌ای است، هرگاه تابع چگالی گسسته‌ی X به‌صورت زیر باشد.

$$f_X(x) = f_X(x;n,p) = \binom{n}{x} p^x q^{1-x} \qquad x = 1, 2, ..., n$$

که در آن $0 \leq p \leq 1$ و $q = 1-p$ و n اعداد صحیح مثبت را اختیار می‌کند.

اگرX دارای توزیع دوجمله‌ای باشد، آنگاه $E(X) = np$ و $Var(X) = npq$.

تعریف ۱-۳-۴. توزیع پواسن

گوییم متغیر تصادفی X دارای توزیع پواسن است هرگاه چگالی X به‌صورت زیر باشد.

$$f_X(x) = f_X(x;\lambda) = \frac{e^{-\lambda}\lambda^x}{x!} \qquad x = 0, 1, 2, ...$$

که $\lambda > 0$.

اگر متغیر تصادفی X دارای توزیع پواسن باشد آنگاه $E(X) = Var(X) = \lambda$.

تعریف ۱-۳-۵. توزیع هندسی

گوییم متغیر تصادفی X دارای توزیع هندسی (پاسکال) است، هرگاه چگالی X به‌صورت زیر باشد.

$$f_X(x) = f_X(x;p) = p(1-p)^x \qquad x = 0, 1, ...$$

اگر X دارای توزیع هندسی باشد آنگاه $E(X) = \frac{q}{p}$ و $Var(X) = \frac{q}{p^2}$.

۱-۳-۲ توزیع‌های پیوسته

تعریف ۱-۳-۶. توزیع یکنواخت یا مستطیلی

یک توزیع بسیار ساده برای یک متغیر تصادفی X به‌صورت زیر می‌باشد.

$$f_X(x) = f_X(x;a,b) = \frac{1}{b-a} \qquad x \in [a,b]$$

که در آن پارامترهای a و b در $-\infty < a < b < \infty$ صدق می‌کنند.

اگرX به‌طور یکنواخت بر $[0,1]$ توزیع شده باشد آنگاه $E(X) = \frac{a+b}{2}$ و $Var(X) = \frac{(b-a)^2}{12}$.

تابع توزیع تجمعی متغیر تصادفی یکنواخت به‌صورت زیر است

$$F_X(x) = \frac{x-a}{b-a} \qquad a < x < b$$

این توزیع الگوی مفیدی را برای چند پدیده‌ی تصادفی فراهم می‌کند. برای مثال اگر بدانیم متغیر تصادفی X تنها در بازه‌ی متناهی مثل $[a,b]$ می‌تواند باشد و اگر فرض کنیم هر دو زیر بازه با طول‌های مساوی از $[a,b]$، با احتمال‌های مساوی X را در بر می‌گیرند، آنگاه X دارای توزیع یکنواخت بر بازه‌ی $[a,b]$ است. وقتی از یک عدد تصادفی از بازه‌ی $[0,1]$ سخن می‌گوییم منظورمان متغیر تصادفی است که به‌طور یکنواخت بر بازه‌ی $[0,1]$ توزیع شده است، می‌باشد.

تعریف ۱-۳-۷. توزیع نرمال

تعداد بسیار زیادی از فنونی که در آمار کاربردی مورد استفاده قرار می‌گیرند بر پایه‌ی توزیع نرمال قرار دارند. گوییم متغیر تصادفی X به‌صورت نرمال توزیع شده است، اگر چگالی آن به‌صورت زیر باشد.

$$f_X(x) = f_X(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad x \in (-\infty, \infty)$$

که $-\infty < \mu < \infty$ و $\sigma > 0$. اگر X متغیر تصادفی نرمال باشدآنگاه $E(X) = \mu$ و $Var(X) = \sigma^2$.

اگر متغیر تصادفی نرمال دارای میانگین ۰ و واریانس ۱ باشد، متغیر تصادفی نرمال استاندارد یا متغیر تصادفی نرمال شده نامیده می‌شود. توزیع نرمال به عنوان الگوی منطقی برای رفتار برخی پدیده‌های تصادفی ظاهر می‌شود. این توزیع شکل حدی بسیاری از توزیع‌های احتمال دیگر نیز هست. توزیع نرمال، توزیع حدی در قضیه‌ی مشهور حد مرکزی نیز می‌باشد.

تعریف ۱-۳-۸. توزیع نمایی

اگر متغیر تصادفیX دارای چگالی به‌صورت زیر باشد، گوییم X دارای توزیع نمایی است.

$$f_X(x) = f_X(x;\lambda) = \lambda e^{-\lambda x} \qquad x \in [0,\infty)$$

که در آن $0 < \lambda$ است. مجموع متغیرهای تصادفی مستقل هم‌توزیع با توزیع نمایی، دارای توزیع گاما است.

اگر X دارای توزیع نمایی باشد آنگاه $E(X) = \frac{1}{\lambda}$ و $Var(X) = \frac{1}{\lambda^2}$.

تعریف ۱-۳-۹. توزیع گاما

گوییم متغیر تصادفی X دارای توزیع گاما است هرگاه دارای چگالی به‌صورت زیر باشد.

$$f_X(x) = f_X(x;r,\lambda) = \frac{\lambda}{\Gamma(r)}(\lambda x)^{r-1}e^{-\lambda x} \qquad x \in [0,\infty)$$

که در آن $0 < r$ و $0 < \lambda$ است.

اگر$r = 1$ باشد، چگالی گاما به چگالی نمایی تبدیل می‌شود. اگرX دارای توزیع گاما با پارامترهای r و باشد آنگاه $E(X) = \frac{r}{\lambda}$ و $Var(X) = \frac{r}{\lambda^2}$.

نکته ۱-۳-۱۰. *توزیع نمایی به عنوان الگویی برای طول عمر چیزهای گوناگون مورد استفاده قرار می‌گیرد. می‌توان نشان داد که طول فاصله‌ی زمانی بین وقایع متوالی، مشروط بر این‌که تعداد وقایع در فاصله‌ی زمانی ثابت دارای توزیع پواسن باشد، دارای توزیع نمایی است. همچنین اگر دوباره فرض کنیم تعداد وقایع در فاصله‌ی زمانی ثابت به‌صورت پواسن توزیع شده، می‌توان نشان داد که طول زمان بین ۰ و لحظه‌ای که واقعه‌ی rام رخ می‌دهد، دارای توزیع گاما است. بنابراین متغیر تصادفی گاما را می‌توان به عنوان متغیر تصادفی زمان انتظار پیوسته تصور کرد که زمانی است که شخص باید برای rامین واقعه منتظر بماند.*