

به نام خدا

تشبیه سازی آماری



(ویرایش اول)

مولفین

رسول طهماسبی

(دانشجوی دکتری دانشگاه صنعتی امیرکبیر)

دکتر صادق رضایی

(هیئت علمی دانشگاه صنعتی امیرکبیر)

شناسنامه کتاب

عنوان :	شبیه سازی آماری
عنوان به انگلیسی :	Statistical Simulation
نام پدیدآورندگان :	صادق رضایی، رسول طهماسبی
نام ناشر :	نشر شار با همکاری انتشارات دانشگاه تفرش
شابک :	978-600-91089-3-0
وضعیت فهرست نویسی :	بر اساس اطلاعات فیبا
رده بندی کنگره :	۱۳۸۹ ش ۲/۵۸/۴۲۷۶ QA
رده بندی دیویی :	۵۱۹/۵
شماره کتابشناسی ملی :	۲۱۶۲۵۴۹
ویراستار علمی و ادبی :	دکتر عادل محمدپور (دانشگاه صنعتی امیرکبیر)
سال انتشار :	چاپ اول اسفند ۱۳۹۰
تیراژ :	۱۰۰۰ نسخه
قیمت :	۷۰۰۰۰ ریال

انتشارات شار



آدرس : تهران، خ انقلاب، خ دانشگاه، خ شهدای ژاندارمری، پ ۵۸ (ساختمان ایرانیان)، واحد ۱۰

تلفن انتشارات : ۴-۶۶۴۸۴۲۹۳

تلفن مرکز بخش : ۶۶۴۰۵۲۲۶ فکس : ۶۶۴۰۵۲۰۸

آدرس سایت فروشگاه اینترنتی : <http://www.eshar.ir>

آدرس سایت اینترنتی انتشارات : <http://www.shar.ir/?Publications>

آدرس پست الکترونیکی : SharPublication@gmail.com



انتشارات دانشگاه تفرش



انتشارات شار

تمام حقوق این اثر برای انتشارات شار محفوظ است

۱۰٪ درآمد سود سهم، موسسه انتشاراتی شار از این سری کتابها متعلق به کانون خیریه فناوریان با مدیریت دکتر سید محسن رزاقی می باشد

فهرست مطالب

۱	۱	شبیه‌سازی
۳	۱	تولید اعداد تصادفی
۴	۱.۱	مقدمه‌ای بر شبیه‌سازی مونت کارلو
۹	۱.۲	تولید اعداد تصادفی یکنواخت
۹	۱.۲.۱	روش Midsquare
۹	۱.۲.۲	روش تولید کننده خطی سازگار
۱۲	۱.۲.۳	تولید اعداد تصادفی یکنواخت در بازه (a, b)
۱۲	۱.۳	روش معکوس یا تبدیل معکوس
۱۳	۱.۳.۱	حالت پیوسته X
۱۵	۱.۳.۲	حالت گسسته X
۱۷	۱.۴	روش قبول-رد
۲۰	۱.۵	تولید اعداد تصادفی برای توزیع‌های مختلف آماری
۲۰	۱.۵.۱	روش Box-Muller برای تولید اعداد تصادفی نرمال
۲۲	۱.۵.۲	روش قطبی برای تولید داده‌های تصادفی نرمال
۲۳	۱.۵.۳	توزیع گاما
۲۴	۱.۵.۴	توزیع کای دو
۲۵	۱.۵.۵	توزیع بتا
۲۶	۱.۵.۶	توزیع دوجمله‌ای
۲۸	۱.۵.۷	توزیع پواسون
۲۸	۱.۶	تمرین‌ها
۳۱	۲	روش مونت کارلو در حل انتگرال‌های معین
۳۲	۲.۱	مقدمه
۳۲	۲.۲	روش Hit-or-Miss

۲۳	روش مونت کارلو میانگین نمونه	۲۰۳
۳۶	روش کنترل تغییرات نمونه‌گیری	۲۰۴
۳۷	روش نمونه‌گیری مهم	۲۰۵
۳۹	روش طبقه‌بندی	۲۰۶
۴۲	روش متغیرهای متقابل	۲۰۷
۴۴	روش برآوردگرهای حداقل مربعات	۲۰۸
۴۸	بحث تکمیلی روش کنترل تغییرات نمونه‌گیری	۲۰۹
۵۱	استفاده از شبیه‌سازی مونت کارلو در روش‌های آماری	۲۱۰
۵۸	تمرین‌ها	۲۱۱
۶۱	انتگرال‌گیری مونت کارلو با استفاده از زنجیر مارکف	۳
۶۲	مقدمه	۳.۱
۶۲	انتگرال‌گیری مونت کارلو	۳.۲
۶۴	انتگرال‌گیری مونت کارلو برای آمار بیزی	۳.۲.۱
۶۵	زنجیر مارکف	۳.۲.۲
۶۶	الگوریتم متروپلیس-هستینگز	۳.۳
۶۷	الگوریتم متروپلیس	۳.۴
۶۹	الگوریتم نمونه‌گیر مستقل	۳.۵
۷۰	الگوریتم متروپلیس-هستینگز تک مولفه‌ای	۳.۶
۷۱	نمونه‌گیری گیز	۳.۷
۷۵	به دست آوردن توزیع‌های شرطی کامل	۳.۷.۱
۷۸	نمونه‌گیری از توزیع‌های شرطی کامل	۳.۷.۲
۷۸	نمونه‌گیری به روش ردی	۳.۸
۷۹	نمونه‌گیری ردی سازگار	۳.۹
۷۹	نمونه‌گیری برشی	۳.۱۰
۸۰	مقدار اولیه و تعیین تعداد تکرارهای سوخته	۳.۱۱
۸۱	مقدار اولیه	۳.۱۱.۱
۸۱	تعیین تعداد تکرارهای سوخته	۳.۱۱.۲
۸۱	تعیین زمان توقف	۳.۱۱.۳
۸۲	روش‌های به دست آوردن نمونه	۳.۱۱.۴
۸۲	تشخیص همگرایی	۳.۱۲
۸۳	آزمون جیوک	۳.۱۲.۱

۸۲	۳.۱۲.۲	روش گلن و روبین
۸۵	۳.۱۲.۳	روش رافتی و لویز
۸۵	۳.۱۳	تمرین‌ها

II استفاده از شبیه‌سازی در روش‌های آماری

۸۹	۴	بازنمونه‌گیری و بوت‌استرپ
۹۰	۴.۱	مقدمه
۹۱	۴.۲	بوت‌استرپ
۹۲	۴.۲.۱	الگوریتم بوت‌استرپ
۹۳	۴.۲.۲	انواع بوت‌استرپ
۹۳	۴.۲.۳	نقاط مثبت و منفی روش بوت‌استرپ
۹۴	۴.۲.۴	بوت‌استرپ در رگرسیون
۹۵	۴.۲.۵	بوت‌استرپ در سری‌های زمانی
۹۵	۴.۲.۶	بوت‌استرپ با داده‌های وابسته
۹۶	۴.۲.۷	آزمون فرض بوت‌استرپ
۹۷	۴.۲.۸	فاصله اطمینان بوت‌استرپ
۱۰۶	۴.۲.۹	آزمون نیکویی برآزش بوت‌استرپ
۱۰۶	۴.۲.۱۰	کاربردهای دیگر بوت‌استرپ
۱۰۶	۴.۳	تکنیک جک‌نایف
۱۰۷	۴.۴	تمرین‌ها
۱۰۹	۵	الگوریتم EM
۱۱۰	۵.۱	مقدمه
۱۱۲	۵.۲	الگوریتم EM
۱۱۲	۵.۲.۱	استخراج الگوریتم EM
۱۱۵	۵.۲.۲	همگرایی الگوریتم EM
۱۱۵	۵.۳	مثال‌های عددی
۱۱۵	۵.۳.۱	توزیع نرمال آمیخته
۱۲۱	۵.۳.۲	برآورد پارامترهای توزیع EL
۱۲۴	۵.۴	تمرین‌ها

۱۲۷	۶	برآورد بیز سلسله مراتبی دو مرحله‌ای
۱۲۸	۶.۱	مقدمه
۱۲۸	۶.۲	استنباط بیزی
۱۲۹	۶.۳	توزیع پیشین و توزیع پسین
۱۳۳	۶.۴	انتخاب توزیع پیشین
۱۳۳	۶.۴.۱	توزیع پیشین فاقد اطلاع
۱۳۴	۶.۴.۲	توزیع‌های مزدوج
۱۳۶	۶.۵	برآورگر بیز
۱۳۸	۶.۶	برآورد نواحی کوچک
۱۴۰	۶.۶.۱	مدل بیز تجربی برای برآورد ناحیه کوچک
۱۴۳	۶.۶.۲	مدل بیز سلسله مراتبی برای برآورد ناحیه کوچک
۱۴۵	۶.۶.۳	آنالیز بیز سلسله مراتبی با استفاده از نمونه‌گیر گیبز
۱۴۷	۶.۷	تمرین‌ها
۱۴۹		پیوست آ: مفاهیم اولیه آمار و احتمال
۱۶۹		پیوست ب: توزیع‌های آماری

لیست تصاویر

۶	نقاط تصادفی در یک مربع به شعاع واحد همراه با یک ربع دایره	۱.۱
۷	سوزن بافون	۱.۲
۳۲	نمودار $g(x)$ در حالت کلی	۲.۱
۵۵	مقایسه نمودار دو تابع	۲.۲
۵۶	نمودار تابع $h(x)$	۲.۳
۵۷	سری زمانی داده‌ها برای بررسی همگرایی	۲.۴
۸۳	سری زمانی دنباله‌های تولید شده برای بررسی همگرایی	۳.۱
۹۱	روش بوت‌استرپ	۴.۱
۱۱۱	نمودار تابع محدب	۵.۱
۱۱۴	نمایش یک مرحله ای الگوریتم EM	۵.۲

پیش‌گفتار

شبیه‌سازی شاخه‌ای از علم آمار است برای پاسخ دادن به سؤالات مربوط به دنیای واقعی، به کمک انجام آزمایش‌هایی مصنوعی که بسیار شبیه وضعیت‌های واقعی هستند. استفاده از شبیه‌سازی مبتنی است بر درک مفاهیم و حل مسائل فراوان و متنوع. دلایل استفاده از شبیه‌سازی متعدد می‌باشند. به عنوان مثال شبیه‌سازی می‌تواند در حل مسائل پیچیده بسیار موثر باشد. در برخی موارد که بنا به دلایل اقتصادی یا محدودیت‌های زمانی، قادر به انجام آزمایش نیستیم، می‌توان از این ابزار قدرتمند آماری بهره جست. گاهی شبیه‌سازی می‌تواند در تایید یا عدم تایید نتایج تحلیلی نیز مورد استفاده قرار گیرد.

از دیدگاهی دیگر شبیه‌سازی تلاشی است برای درک رفتار فرآیندهای پیچیده تحت شرایط متغیر. برای مثال در فرآیند طراحی دستگاه‌های هدایت‌کننده الکترونیکی سفینه‌های فضایی، دانشمندان علاقمند هستند که در طراحی‌های مختلف، احتمال از کار افتادن دستگاه را حساب کنند. یک راه محاسبه این احتمال این است که دستگاه را بسازند و آن را در شرایط واقعی قرار دهند و امتحان کنند؛ ولی این کار از لحاظ کمیت زمانی و هزینه مقرون به صرفه نمی‌باشد. روش دیگر این است که عمل دستگاه هدایت‌کننده الکترونیکی را روی رایانه شبیه‌سازی کنند. به این ترتیب می‌توان دستگاه را تحت شرایط آزمایشی گوناگون مشاهده و احتمال‌های از کار افتادن را به خوبی محاسبه کرد. در اینجا شبیه‌سازی به محققین کمک می‌کند تا طرحی را برگزینند که در آن احتمال از کار افتادن دستگاه کمترین مقدار باشد.

عدم وجود مرجعی کامل که در برگزیده مفاهیم و روش‌های گوناگون شبیه‌سازی آماری باشد از یک طرف و عدم وجود منبعی فارسی از طرف دیگر، موجب تالیف این کتاب گردید. کتاب حاضر نتایج تجربیات چندین ساله مولفین می‌باشد که پس از اعمال تغییرات زیادی در نسخه اولیه تحت عنوان «شبیه‌سازی آماری» ارائه گردیده است. سعی شده شیوه نگارش این کتاب ساده و روان و همراه با مثال‌های متعدد باشد. در بسیاری از موارد برنامه حل مثال‌ها به زبان MATLAB و $S++$ ارائه شده است. این کتاب می‌تواند به عنوان مرجعی مناسب برای دانشجویان مقاطع کارشناسی و کارشناسی ارشد در رشته‌های مهندسی و علوم پایه مورد استفاده قرار گیرد. با این وجود واقف هستیم که مجموعه حاضر خالی از ایراد و اشکال نیست و لذا هر نوع یادآوری و راهنمایی برای بهتر شدن این کتاب مورد استقبال مولفین قرار خواهد گرفت.

کتاب پیش رو به دو بخش تقسیم شده است. بخش اول که شامل سه فصل می‌باشد، مربوط به شبیه‌سازی آماری می‌باشد. بخش دوم استفاده از شبیه‌سازی در روش‌های آماری است که نیازمند دانش آماری بیشتری می‌باشد.

در فصل اول شیوه‌های تولید اعداد تصادفی را با روش‌های ساده، که پایه تولید اعداد تصادفی از دیگر توزیع‌های آماری می‌باشند، ارائه کرده‌ایم. در ادامه این فصل روش‌های تولید اعداد تصادفی برای دیگر توزیع‌های آماری را نیز ارائه کرده‌ایم. روش‌های گوناگون شبیه‌سازی مونت کارلو برای حل انتگرال‌های معین، در فصل دوم بیان گردیده است. در فصل سوم انتگرال گیری مونت کارلو را با استفاده از زنجیر مارکف (MCMC) تشریح کرده‌ایم و الگوریتم‌های متروپلیس-هستینگز و گیبز را نیز معرفی کرده‌ایم. در انتهای این فصل کاربرد MCMC را در روش‌های آماری بیان کرده‌ایم.

روش‌های باز نمونه‌گیری را که شامل بوت‌استرپ و جک‌نایف می‌باشد، در فصل چهارم ارائه کرده‌ایم و در فصل پنجم الگوریتم EM را برای برآورد حداکثر درست‌نمایی پارامترهای مدل با ذکر مثال‌های واقعی و کاربردی تشریح کرده‌ایم. فصل آخر را به دلیل نیازمند بودن به مفاهیم آمار بیز، به بررسی برآورد بیز سلسله مراتبی دو مرحله‌ای اختصاص داده‌ایم. قسمت آخر این فصل را به بررسی یک مفهوم کاربردی در شبیه‌سازی توزیع پسین برای برآورد نواحی کوچک اختصاص دادیم.

در پیوست آ مفاهیم اولیه آمار و احتمال را تعریف کرده‌ایم که به منظور یادآوری، آمادگی و امکان دور از دسترس بودن منابع این‌ها تنظیم شده است و به اختصار به یادآوری نکات مباحث احتمال، اشاره می‌کنیم. همچنین در پیوست ب به اختصاص تابع احتمال توزیع‌های گسسته و تابع چگالی احتمال توزیع‌های پیوسته را معرفی و برخی از مشخصه‌های آن‌ها را همچون میانگین و واریانس، ذکر کرده‌ایم.

در طول کتاب سعی شده برنامه حل بیشتر مثال‌ها ارائه شوند. دانشجویان می‌توانند برنامه‌ها را از طریق آدرس www.rnsp.org/StatisticalSimulation/ دریافت کنند. توصیه می‌شود که برنامه‌ها توسط دانشجویان مجدداً اجرا شود و با ایجاد تغییرات در آن‌ها، مهارت برنامه نویسی خود را افزایش دهند. تمرین‌ها طوری طراحی و انتخاب شده‌اند که بتواند به درک بیشتر مطالب ذکر شده در هر فصل کمک نمایند. حل تمرین‌ها می‌تواند دانشجویان را در پیشبرد مهارت‌های برنامه نویسی یاری رساند.

نماد گذاری

a.s.	almost surly
$B(x, y)$	تابع بتا
$BB(n_\gamma, n_\tau, n_\pi)$	توزیع بتا-دو جمله‌ای
$Beta(\alpha, \beta)$	توزیع بتا
$Bin(n, p)$	توزیع دو جمله‌ای
$Can(\theta, \sigma)$	توزیع کوشی
$cf(t)$	تابع مشخصه
$Corr(X, Y)$	ضریب همبستگی
$Cov(X, Y)$	کوواریانس
$CV(.)$	ضریب تغییرات
$Dir_k(\alpha_0, \alpha_1, \dots, \alpha_k)$	توزیع درخله
$DU(\{\circ, \backslash, \dots, \theta\})$	توزیع گسسته یکنواخت
$DWei(\beta, p)$	توزیع وایبل گسسته
$\mathbb{E}(X)$	میانگین
$EV(\alpha, \beta)$	توزیع مقادیر غایی
$Exp(\lambda)$	توزیع نمایی
$F(n_\gamma, n_\tau)$	توزیع F
${}_pF_q(.)$	تابع هایپرژئومتریک
$Gam(\alpha, \lambda)$	توزیع گاما
$Geo(p)$	توزیع هندسی
$GHG(N, M_\gamma, M_\tau; n)$	توزیع فوق هندسی تعمیم یافته
$GPar(\alpha, \lambda)$	توزیع پاراتو تعمیم یافته
$HG(N, m, n)$	توزیع فوق هندسی
$IG(\mu, \sigma^2)$	توزیع گاوسی معکوس
$IGam(\mu, \lambda)$	توزیع گامای معکوس

$\text{Lap}(\mu, \beta_1, \beta_2)$	توزیع لاپلاس
$\text{LCG}(\cdot)$	تولید کننده خطی سازگار
$\text{Lo}(\alpha, \beta)$	توزیع لجستیک
$\text{LogN}(\mu, \sigma^2)$	توزیع لاگ نرمال
$\text{LS}(\theta)$	توزیع سری های لگاریتمی
$\text{Med}(X)$	میانه
$\text{mgf}(t)$	تابع مولد گشتاور
MLE	برآورد ماکسیمم درستنمایی
$\text{Mod}(X)$	مد
MSE	میانگین مربع خطا
$\text{Mul}(X_1, \dots, X_k)$	توزیع چند جمله ای
$N(\mu, \sigma^2)$	توزیع نرمال
$N_p(\mu, \Sigma)$	توزیع نرمال p بعدی
$\text{NB}(n, p)$	توزیع دو جمله ای منفی
$\mathbb{P}(X = x)$	احتمال
$\text{Par}(\alpha, \beta)$	توزیع پارانو
$\text{pgf}(t)$	تابع مولد احتمال
$\text{Poi}(\lambda)$	توزیع پواسون
$\mathbb{R}$	مجموعه اعداد حقیقی
$t_{(\alpha)}$	توزیع تی
$\text{TExp}(\mu, \lambda)$	توزیع نمایی بریده شده
$\text{Tri}(a, b, m)$	توزیع مثلثی
$U(a, b)$	توزیع یکنواخت
$\text{Var}(X)$	واریانس
$\text{Wei}(\alpha, \lambda)$	توزیع وایبل
$\mathbb{Z}$	مجموعه اعداد صحیح
$\Gamma(\alpha)$	تابع گاما
$\gamma_1(\cdot)$	ضریب چولگی
$\gamma_2(\cdot)$	ضریب برجستگی
χ^2_r	توزیع کای دو