شبیه سازی و تولید اعداد تصادفی

(R)

تدوین: مرکز تحلیل آماری خوارزمی

❋❋❋

www.kharazmi-statistics.ir

مرکز آماری خوارزمی

مقدمه:

در بسیاری از سیستم‌ها برای تجزیه و تحلیل مسائل مختلف آن و یا اصولا جهت طراحی، استقرار و توسعه سیستم‌ها، تکنیک‌هایی لازم است که بتوان حالت‌ها و شرایط خاص و حالات تصادفی سیستم را نیز مورد بررسی و مطالعه قرار داد. شبیه‌سازی یکی از کارآمدترین تکنیک‌های شناخته شده است.

شبیه‌سازی قادر است سیستم مورد مطالعه را تحت شرایط خاص و بحرانی و احتمالی که سیستم با آن مواجه است بررسی و تجزیه و تحلیل کند. این قابلیت شبیه‌سازی شدیدا به اعداد تصادفی وابسته است. در واقع پایه اصلی توانمندی شبیه سازی بر اعداد تصادفی استوار است.

هرچه تولید اعداد تصادفی با معیارهای تصادفی بودن سازگارتر باشد، بالطبع تاثیر بسزایی در جهت اثر بخشی مدلسازی و تجزیه و تحلیل نتایج شبیه‌سازی خواهد داشت. در این متن تلاش شده است که روش‌های تولید اعداد تصادفی برای توزیع‌های مختلف در نرم افزار R آموزش داده شود.

مطالبی که در این متن بیان شده است عبارتنداز:

اعداد تصادفی و کاربردهای آن

تولید اعداد تصادفی با توزیع یکنواخت

تولید اعداد تصادفی با توزیع برنولی و دوجمله‌ای

تولید اعداد تصادفی با توزیع پواسن

تولید اعداد تصادفی با توزیع نمایی

تولید اعداد تصادفی با توزیع نرمال

اعداد تصادفی و کاربردهای آن

عدد تصادفی، عدد تولید شده از یک فرایند است که نتیجه آن غیر قابل پیش‌بینی است. در هر بار نوشتن فرمان تولید اعداد تصادفی در نرم‌افزار اعداد جدید تولید می‌شود که با اعداد تولید شده در مرتبه‌ی قبل متفاوت است.

کاربردها:

شبیه‌سازی: وقتی یک کامپیوتر برای شبیه‌سازی مفاهیم طبیعی مورد استفاده قرار می‌گیرد، اعداد تصادفی برای واقعی نشان دادن اجزا و رویدادها مورد نیاز هستند.

نمونه‌برداری: تست کردن همه حالات ممکن برای یک سیستم اغلب غیر عملی است اما وضعیت و درستی یک نمونه تصادفی می‌تواند حالت کلی سیستم را شرح دهد.

آنالیز عددی: روش‌های مبتکرانه‌ای برای حل مسائل عددی پیچیده ابداع شده است که از اعداد تصادفی استفاده می‌کنند.

برنامه‌نویسی کامپیوتری: مقادیر تصادفی منابع خوبی از اطلاعات برای تست کردن کارایی الگوریتم‌های کامپیوتری هستند؛ از همه مهم‌ترین نقش آنها در اجرای الگوریتم‌های تصادفی است .

تصمیم‌گیری: گزارش‌هایی مبنی بر اینکه بعضی مدیران اجرایی تصمیمات خود را بر پایه پرتاب سکه و یا دارت می‌گیرند؛ در واقع بعضی وقت‌ها باید بدون غرض‌ورزی تصمیمات گرفته شوند.

تولید اعداد تصادفی با توزیع یکنواخت

دستور کلی تولید اعداد تصادفی با توزیع یکنواخت به صورت زیر است.

>runif (n, min=a, max=b)

مثال: برای تولید ۶ عدد تصادفی با توزیع یکنواخت بین اعداد [۰و۱] در نرم افزار دستور زیر را بنویسید.

>runif(۶)

همانطور که در خروجی نرم‌افزار مشاهده می‌کنید با دوبار نوشتن فرمان مورد نظر اعداد داده شده متفاوتی تولید شده است و با هربار تکرار این دستور اعداد تولید شده متفاوت می‌شود.

```
> runif(6)
[1] 0.4479037 0.5419296 0.1188486 0.4828799 0.9699076 0.2709823
> runif(6)
[1] 0.39969690 0.03041531 0.68554379 0.66458973 0.07711407 0.49575033
```

مثال: ۲۰ عدد تصادفی بین بازه ی ۵- و ۷ با توزیع یکنواخت تولید کنید.

<runif(۲۰, min=-۵, max=۷)

```
> runif(20, min=-5, max=7)
 [1]  3.1091182  5.7440470  4.0226651  3.1344581  0.3064481  0.7736489
 [7]  2.6352325 -3.8893652 -1.8448601  5.5058267  1.3215572 -2.4078605
[13]  0.5930902  6.1376590 -4.6311579 -3.2354417  1.1721841 -4.0283137
[19] -0.5279936 -0.2216779
```

تولید اعداد تصادفی با توزیع برنولی و دوجمله‌ای

آزمایش برنولی روشی است که در آن تنها دو نتیجه به دست می‌آید. مانند سکه که نتایج شیر یا خط دارد و یا سالم و معیوب بودن لامپ. در واقع آزمایش برنولی برآمد موفقیت و شکست است.

فرض کنید منظور از x، جمع m متغیر تصادفی مستقل برنولی باشد که هر کدام احتمال p را می‌گیرند و آنگاه x یک متغیر تصادفی دو جمله‌ای نامیده می‌شود که بیانگر تعداد موفقیت‌ها در m آزمایش برنولی است. m متغیر تصادفی برنولی مجموعه‌ی$\{۰,۱,۲,...,m\}$را شامل می‌شود. (m=۱ توزیع برنولی محسوب می‌شود) احتمال یک متغیر تصادفی x به ازای هر کدام از مقادیر فوق توسط توزیع دو جمله‌ای کنترل می‌شود:

$$P(X = x) = \binom{m}{x} p^x (۱ - p)^{m-x}$$

دستور کلی تولید اعداد تصادفی با توزیع دوجمله‌ای به صورت زیر است:

>rbinom(n , size=m, prob=b)

مثال: فرض کنید ۱۵٪ تیوب‌های تولید شده توسط یک دستگاه خراب است. با فرض اینکه این تیوب‌ها مستقل از هم تولید شوند و در هر ساعت ۲۵ تیوب توسط این دستگاه تولید شود، اگر در هر ساعت بیش از چهار تیوب دچار خرابی باشند، این فرآیند تولید خارج از کنترل تلقی می‌شود. تعداد تیوب‌های معیوب تولید شده در هر ساعت توسط این ماشین را در طول یک شبانه‌روز شبیه‌سازی کنید و با توجه به آن تعیین کنید که آیا فرآیند ساخت تیوب قابل کنترل است یا نه؟

حل: از آنجا که ۲۵ تیوب در هر ساعت تولید می‌شود و شانس معیوب بودن در هر کدام ۰/۱۵ است، مستقل از تیوب‌های دیگر، تعداد تولیدات در هر یک ساعت یک متغیر تصادفی دو جمله‌ای است با m=۲۵ و p=۰/۱۵. برای شبیه‌سازی تعداد معیوب‌ها برای یک ساعت در طول شبانه‌روز لازم است ۲۴ داده‌ی تصادفی دوجمله‌ای تولید شود. سپس تعداد نمونه‌هایی که خرابی‌های آن را از ۴ تجاوز می‌کند را شناسایی می‌کنیم. دستورات را در نرم‌افزار به صورت زیر بنویسید.

```
> a<-rbinom(۲۴,۲۵,۰.۱۵)
>a
>any(a>۴)
```

خروجی مورد نظر در نرم‌افزار به صورت روبرو است. همانطور که مشاهده می‌کنید. طبق شبیه‌سازی انجام گرفته ساعاتی از شبانه‌روز وجود دارد که تولیدات معیوب در آن ساعات بیش از ۴ عدد است. در نتیجه تولید تیوب در این خط تولید کنترل شده نیست.

```
> a
 [1] 2 3 4 6 5 4 3 4 4 1 3 6 4 2 3 6 1 2 4 1 3 3 5 2
> any(a>4)
[1] TRUE
```

برای محاسبه مقدار احتمال $P(X=x)$ دستور زیر نوشته می‌شود.

dbinom(x , size=m, prob=b)

عبارت روبرو را در نظر بگیرید. $P(X=۷)=\binom{۹}{۷}۰.۲^{۷}(۱-۰.۲)^{۲}$

برای محاسبه‌ی احتمال روبرو می‌توان به صورت زیر عمل کرد.

```
> dbinom(7,9, 0.2)
[1] 0.000294912
```

برای محاسبه‌ی احتمال از دستور زیر استفاده می‌شود. $P(X\leq ۷)$

```
> pbinom(7,9, 0.2)
[1] 0.9999811
```

برای محاسبه‌ی چندک در توزیع دوجمله‌ای نیز می‌توانید از دستور زیر استفاده کنید. در اینجا چندک ۵۹ در نظر رفته شده است.

```
> qbinom(0.59,9, 0.2)
[1] 2
```

تولید اعداد تصادفی با توزیع پواسن

توزیع پواسن، حد دنباله‌ای از توزیع دو جمله‌ای با پارامترهای n و p_n است وقتی که n به سمت بی‌نهایت و p_n به سمت صفر میل می‌کند در حالی که میانگین np_n به مقدار ثابت λ همگراست. واریانس $np_n(۱-p_n)$ نیز به همان مقدار λ همگراست. بنابراین میانگین و واریانس یک متغیر تصادفی یک متغیر تصادفی پواسن هر دو برابر با λ است.

وقتی طول زمان به n بازه‌ی مستقل تقسیم می‌شود به طوری که در هر بازه مقادیر صفر یا ۱ گنجانده شود، متغیر تصادفی پواسن شمارشگر کل اعداد درون بازه‌هاست. نمونه‌ای از داده‌ها با توزیع پواسن در طبیعت مانند تعداد زلزله‌هایی که در یک ناحیه در طول یک سال رخ می‌دهد یا تعداد افرادی که در یک ساعت به یک صندوق‌دار بانک مراجعه می‌کنند، است.

مقادیر ممکنی که متغیر تصادفی پواسن می‌تواند بگیرد در مجموعه‌ی اعداد صحیح غیرصفر {۰,۱,۲,...}و احتمال پذیرفتن هر کدام از این مقادیر برابر است با:

$$P(X = x) = \frac{exp^{-\lambda}\lambda^{x}}{x!} \; , x = ۰,۱,۲, ...$$

دستور کلی برای تولید اعداد تصادفی با توزیع پواسن به صورت زیر است.

> rpois(n , lambda)

در اینجا lambda پارامتر نسبت پواسن است زمانی که n تعداد مقادیر تولید شده است.

مثال: تعداد تصادفات سالیانه را در طول ۱۵ سال برای حالتی که ۲.۸ تصادف در سال رخ می‌دهد با فرض مدل پواسن برای تعداد تصادفات سالیانه شبیه‌سازی کنید.

برای شبیه‌سازی مربوطه باید دستور زیر را در نرم‌افزار نوشت.

> rpois(۱۵,۲.۸)

نتیجه به صورت روبرو است.

```
> rpois(15,2.8)
 [1] 5 2 4 4 3 3 1 4 5 2 2 3 2 2 4
```

- با دستور dpois(x,lambda) مقدار احتمال محاسبه می شود. که در آن x تعداد رخدادهای پواسن است.دستور ppois(x,lambda) مقدار فراوانی تجمعی $P(X \le x)$ و qpois(۰.۶۳,lambda) چندک‌های پواسن (در اینجا صدک ۶۳ام) را محاسبه می‌کند.

تولید اعداد تصادفی با توزیع نمایی

متغیرهای تصادفی نمایی به عنوان مدل ساده‌ای برای تخمین تعداد دفعات شکست به کار گرفته می‌شود که در بخش‌های مکانیکی و الکترونیکی نیز مورد استفاده قرار می‌گیرد و معمولا بیانگر زمانی است که طول می‌کشد تا

یک فرآیند به موفقیت برسد یا به صورت کامل انجام پذیرد و توزیع نمایی توسط میزان شکست‌ها (λ) نمایش داده می‌شود.

T دارای توزیع نمایی با میانگین $\lambda > 0$ است اگر به ازای هر t داشته باشیم.

$$P(T \geq t) = ۱ - e^{-\lambda t}$$

شکل کلی دستور تولید اعداد تصادفی در نرم‌افزارR به صورت زیر است.

>rexp(n,rate) n تعداد متغیر تولید شده و rate پارامتر توزیع

خروجی نرم‌افزار برای ۱۰ عدد تصادفی تولید شده و پارامتر ۳

```
> rexp(10,3)
 [1] 0.31446354 0.30189120 0.60367863 1.37246832 0.09740214 0.12918150
 [7] 1.35077568 0.10871274 0.75202392 0.08306590
```

- با دستور dexp(t,lambda) مقدار احتمال محاسبه می‌شود. که در آن t تعداد رخدادهای پواسن است. دستور pexp(q,rate) مقدار فراوانی تجمعی $P(T \leq q)$ و qexp(۰.۴۸,lambda) چندک-های پواسن (در اینجا صدک ۴۸ام) را محاسبه می‌کند.

تولید اعداد تصادفی با توزیع نرمال

متغیر تصادفی نرمال x دارای تابع چگالی احتمال $f(x)$ است به طوری که

$$f(x) = \frac{۱}{\sigma\sqrt{۲\pi}} e^{-\frac{(x-\mu)^۲}{۲\sigma^۲}}$$

که در آن μ و $\sigma^۲$ به ترتیب میانگین و واریانس x می‌باشد. متغیر تصادفی نرمال استاندارد دارای میانگین $\mu = ۰$ و انحراف استاندارد $\sigma = ۱$ است.

دستور کلی تولید اعداد تصادفی با توزیع نرمال به صورت زیر است.

>rnorm(n,mean,sd)

در آن n تعداد متغیر تولید شده، mean میانگین توزیع و sd انحراف از میانگین است.

خروجی نمایش داده شده برای نرم‌افزار ۱۰ عدد تصادفی با توزیع نرمال، با میانگین ۳ و انحراف معیار ۰.۴ است.

```
> rnorm(10,3,0.4)
 [1] 2.776120 3.946311 2.520366 3.473714 3.568528 2.536124 2.295184 3.020533
 [9] 3.659661 3.629980
```

دستور dnorm() مقدار احتمال، pnorm() مقدار فراوانی تجمعی و qnorm(۰.۳۲,mean,sd) چندک‌های پواسن (در اینجا صدک ۳۲ام) را محاسبه می‌کند.

منبع:

- http://www.mortezamansoori.blogfa.com/post-۲۲۲.aspx

- نقش اعداد تصادفی در شبیه سازی و بررسی تحلیل الگوریتم‌های تولید اعداد تصادفی و ارائه روش تلفیقی جدید،عبداله آقایی، مهدی زاهدی شولمی، نشریه بین المللی مهندسی صنایع و مدیریت تولید دانشگاه علم و صنعت ایران، جلد ۱۹، شماره ۴، زمستان ۱۳۸۷، صفحه ۲۷-۱۷

- مبانی برنامه نویسی آماری با R، جان براون و دانکن مرداک، ترجمه دکتر منوچهر بابانژاد و مبین ملکشاه، انتشارات دانشگاه گلستان،۱۳۹۳