אקטואר מחפש סימולציה על החוף של מונקו

מדען הנתונים המוערך האקטואר רועי פולניצר מסביר במאמר זה כי סימולציית מונטה קרלו היא טכניקה מתמטית, המסייעת בקבלת החלטה בתנאי אי וודאות המשלבת שימוש במספרים אקראיים, בהתפלגויות ערכים אפשריות ובהסתברויות, ובכך למעשה עוזרת בפתרון בעיות הקשורות לניהול סיכונים ותחזיות לעתיד.

Oימולציית מונטה קרלו, הקרויה על שם בירת ההימורים המפורסמת של הנסיכות הקטנה מונקו, הינה מתודולוגיה אקטוארית חזקה מאוד. עבור האקטואר, סימולציה פותחת צוהר לפתרון בעיות קשות ומורכבות בקלות רבה.

מונטה קרלו יוצרת עתיד מלאכותי על ידי דגימה מקרית של אלפי ואף מיליוני מסלולים (תרחישים או תסריטים) דמיוניים (מתורחשים או מסומלצים) של תוצאות ומסתכלת על המאפיינים הסטטיסטיים שלהם. לשכת מעריכי השווי והאקטוארים הפיננסיים בישראל בראשותי סבורה שאקטוארים לא צריכים ללמוד קורסים מתקדמים במתמטיקה כמו אלה שאותם למדתי לתואר שני בכלכלה כי זה פשוט לא הגיוני ולא מעשי.

אקטואר מבריק יעשה שימוש בכל הכלים הזמינים לו כדי לקבל את אותה תשובה בדרך הקלה והפרקטית ביותר. ובכל המקרים, כאשר ממדלים את סימולציית מונטה קרלו בצורה נכונה, היא מספקת תשובות הדומות לאלו המתקבלות משיטות אלגנטיות יותר מבחינה מתמטית.

אז מהי אם כן סימולציית מונטה קרלו וכיצד היא עובדת?

מונטה קרלו או סנטו מרינו?

סימולציית מונטה קרלו (שבמשך השנים לקוחות לא מעטים שלי קראו לה בגיחוך "סימולציית סנטו מרינו") בצורתה הפשוטה הינה מחולל מספרים אקראיים המשמש לניבוי/חיזוי, הערכה/אמידה וניתוח/ניהול סיכונים. סימולציה כשלעצמה מריצה תרחישים רבים הנגזרים ממודל מסוים על ידי בחירה חוזרת של ערכים מתוך ***התפלגות*** שהוגדרה מראש על ידי האקטואר עבור המשתנים הלא וודאיים ושימוש בערכים אלה עבור המודל.

כל התרחישים הללו מייצרים תוצאות הקשורות למודל, שבו לכל תרחיש יכולה להיות **תחזית**. תחזיות הן מאורעות (בדרך כלל עם נוסחאות או פונקציות) שאותן האקטואר מגדיר בתור ***משתני יציאה*** חשובים של המודל.

מאורעות אלו יכולים להיות סכומים, רווח נקי או הוצאות. באופן פשטני, ניתן לחשוב על גישת סימולציית מונטה קרלו כעל הגרלת כדורי לוטו מתוך מכונת הגרלה גדולה שוב ושוב עם החזרה של הכדורים (כלומר, מגרילים באקראיות כדור לוטו מסוים מתוך מכונת ההגרלה ולאחר מכן מחזירים אותו פנימה למכונה).

גודלה וצורתה של מכונת ההגרלה תלויים בהנחת ההתפלגות בתור ***משתנה כניסה*** (למשל, התפלגות נורמלית עם תוחלת של 100 וסטיית תקן של 10, לעומת מרחב התפלגות אחיד או התפלגות טריאנגולרית) כאשר קיימות מכונות הגרלה גדולות יותר או בעלות צורה שונה ממכונות הגרלה אחרות, תכונות אשר מאפשרות להגריל מתוכן כדורים בתדירות גבוהה יותר על פני מכונות הגרלה אחרות. מספר כדורי הלוטו המוגרלים שוב ושוב תלוי במספר ***הניסויים*** (ניסיונות) המתורחשים.

אם ניקח למשל, מודל מורכב הנשען על הנחות רבות, הרי שניתן לדמות את אותו מודל מורכב למכונת הגרלה גדולה מאוד, אשר בתוכה שוכבות הרבה מאוד מכונות הגרלה קטנטנות. לכל מכונת הגרלה קטנטנה יש סט כדורי לוטו משלה המקפצים בתוכה.

לפעמים מכונות ההגרלה הקטנטנות הללו מתואמות האחת עם השנייה (אם יש **קורלציה** בין המשתנים) וכדורי הלוטו שלהן קופצים במקביל בעוד שכדורי לוטו אחרים קופצים ללא תלות זה בזה. כדורי הלטו שמוגרלים בכל פעם מהאינטראקציות הללו בתוך המודל המורכב (מכונת ההגרלה הגדולה) מוצגים בטבלה ומתועדים, ובכך למעשה מספקים את ***התוצאה החזויה*** של הסימולציה.

זרע אקראי

חשוב לציין שסימולציה מעצם הגדרתה תניב בכל הרצה שלה (קרי, בכל פעם שהיא מורצת) תוצאות מעט שונות – זה גם מה שקורה לנו בכל פעם שאנו מריצים אותה שאלה ב- Chat GPT. תכונה זו נגזרת מרוטינת דגימת המספרים האקראיים בסימולציית מונטה קרלו והיא עובדה תיאורטית בכל מחוללי המספרים האקראיים.

בעת פרזנטציה, לעיתים האקטואר רוצה שאותן תוצאות שהוצגו בדו"ח האקטוארי שלו יופיעו גם במהלך הפרזנטציה, או לחילופין כאשר הוא משתף אקסלים עם אחרים אז הוא רוצה שיתקבלו אותן תוצאות בכל פעם. לשם כך, בתוכנות סטטיסטיות מסוימות כולל בשפת R ושפת פייתון (שפות הפיתוח השולטות בעולמות ה- Data והאקטואריה) יש פיצ'ר הנקרא "זרע אקראי" (Random Seed).

זרע אקראי הינו מספר שלם חיובי שתפקידו להבטיח שעבור אותו מספר של ניסויים ואותם משתני כניסה, הסימולציה תניב תמיד את אותו רצף של מספרים אקראיים, ובכך תבטיח את קבלתו של אותו סט תוצאות סופי.

בקרת דיוק

כלי חזק מאוד בסימולציית מונטה קרלו הוא בקרת דיוק (Precision Control). למשל, נשאלת השאלה, כמה ניסויים נחשבים למספיקים כדי להריץ מודל מורכב? בקרת הדיוק מייתרת הלכה ולמעשה את מלאכת ניחוש מספר הניסויים הרלוונטי בכך שהיא מאפשרת לסימולציה לעצור ברגע שמגיעים לרמת הדיוק שנקבעה מראש.

הפונקציונליות של בקרת הדיוק מאפשרת לאקטואר להגדיר את מידת הדיוק שאליה הוא רוצה להגיע עבור התחזית שלו. באופן כללי, ככל שמריצים יותר ניסויים, כך רווח הסמך (Confidence Interval, רווח בר-סמך) מצטמצם והסטטיסטיקה נעשית מדויקת יותר.

שגיאה, דיוק וביטחון

תכונת בקרת הדיוק משתמשת במאפיין של רווחי סמך כדי לקבוע מתי הושגה רמת הדיוק המוגדרת עבור סטטיסטי מסוים (כגון: תוחלת, סטיית תקן, א-סימטריה, גבנוניות וכו'). עבור כל תחזית, האקטואר יכול להגדיר את רווח הסמך הספציפי שאותו הוא רוצה עבור רמת הדיוק.

חשוב מאוד שלא להתבלבל בין שלושה מונחים שונים מאוד: ***שגיאה, דיוק ו- ביטחון*** (סמך). למרות שהם נשמעים דומים, המושגים הללו שונים זה מזה באופן משמעותי.

כאשר אני מלמד אקטואריה אני נוהג להיעזר בדוגמא מוחשית כדי להסביר את שלושת המושגים הללו. נניח שאני יצרן ביסקוויטים ושאני מעוניין לדעת כמה ביסקוויטים שבורים ישנם בממוצע בכל חבילה של 100 ביסקוויטים. אחת הדרכים לעשות זאת היא לאסוף מדגם של חבילות ארוזות מראש בנות 100 ביסקוויטים כל אחת, לפתוח אותן ולספור כמה ביסקוויטים שבורים יש בכל אחת מהן.

נניח שאני מייצר מיליון חבילות ביסקוויטים כאלו ביום (זוהי למעשה ***האוכלוסייה*** שלי) אבל אני פותח באקראי רק 10 חבילות ביסקוויטים (זהו גודל ***המדגם*** שלי, המכונה גם מספר ***הניסויים*** שלי בסימולציה). נניח שמספר הביסקוויטים השבורים בכל חבילת ביסקוויטים הוא כדלקמן: 24, 22, 4, 15, 33, 32, 4, 1, 45 ו-2.

לפיכך, המספר הממוצע של הביסקוויטים השבורים בכל חבילת ביסקוויטים 18.2. בהתבסס על 10 דגימות או ניסויים אלו, הממוצע הוא 18.2 ביסקוויטים שבורים בחבילת ביסקוויטים, בעוד שבהתבסס על המדגם, רווח הסמך (confidence interval, רווח בר-סמך) בהסתברות של 80% מספר הביסקוויטים השבורים בחבילת ביסקוויטים נע בין 2 ביסקוויטים שבורים ל- 33 ביסקוויטים שבורים.

במילים אחרות, ב- 80 אחוז מהמקרים מספר הביסקוויטים השבורים בחבילה יהיה איפשהו בין 2 ביסקוויטים שבורים ל- 33 ביסקוויטים שבורים, וזאת כאמור ***בהתבסס על גודל מדגם זה או על מספר הניסויים שהורצו.*** יחד עם זאת, נשאלת השאלה עד כמה אני באמת בטוח ש- 18.2 הוא הממוצע הנכון של הביסקוויטים השבורים בכל חבילת ביסקוויטים?

אבל האם באמת מספיקים לי רק 10 ניסויים כדי לקבוע זאת? רווח הסמך שקיבלתי בין 2 ביסקוויטים שבורים ל- 33 ביסקוויטים שבורים הוא רחב מדי ולא פרקטי. נניח שאני רוצה ערך ממוצע מדויק יותר שבו השגיאה היא ±2 (פלוס/מינוס 2) ביסקוויטים שבורים ב-90 אחוז מהזמן.

במילים אחרות, זה אומר שאם אני אפתח את ***כל*** מיליון חבילות הביסקוויטים שאני מייצר ביום, רק ב- 900,000 מהן יהיו בממוצע ביסקוויטים שבורים במספר ביסקוויטים תוחלתי מסוים ($\bar{x}$) פלוס/מינוס 2 ביסקוויטים שבורים. נשאת השאלה, כמה חבילות ביסקוויטים נוספות יהיה עליי לדגום כדי להגיע לרמת דיוק כזו?

שימו לב שבמקרה דנן שלפנינו, 2 ביסקוויטים שבורים הם רמת השגיאה, בעוד ש- 90 האחוז הם רמת הדיוק. אם נריץ מספר מספיק של ניסויים, אזי רווח הסמך של 90 אחוז יהיה זהה לרמת הדיוק של 90 אחוז, כאשר מדד מדויק יותר עבור הממוצע מתקבל כך שב- 90 אחוז מהזמן, השגיאה, ומכאן, הביטחון יהיה ±2.

לשם הדוגמה, נניח שהממוצע הוא 20 ביסקוויטים שבורים, ואז רווח הסמך של 90 אחוז יהיה בין 18 ל-22 ביסקוויטים שבורים, כאשר רווח סמך זה מדויק ב- 90 אחוז מהזמן, בעוד שבעת בפתיחת כל מיליון חבילות הביסקוויטים שאני מייצר, ב-900,000 מהן יהיו בין 18 ביסקוויטים שבורים ל- 22 ביסקוויטים שבורים. מספר הניסויים הנדרש כדי להגיע לדיוק שכזה מבוסס על משוואת שגיאת הדגימה הבאה:

$$\bar{x} \pm Z\frac{s}{\sqrt{n}}$$

כאשר $Z\frac{s}{\sqrt{n}}$ היא השגיאה של 2 ביסקוויטים שבורים, $\bar{x}$ הוא ממוצע המדגם, Z הוא ציון התקן הנורמלי סטנדרטי (אם תרצו: מספר סטיות התקן) המתקבל מרמת דיוק של 90%, s היא סטיית התקן של המדגם ו- n הוא מספר הניסויים הנדרש כדי להגיע לרמת שגיאה זו ברמת הדיוק שהוגדרה.

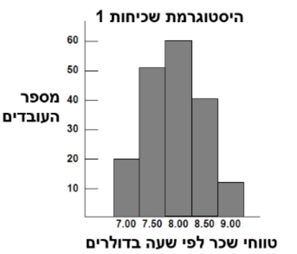
התפלגויות

לפני שצוללים לתוך סימולציית מונטה קרלו, צריך ראשית להבין את הרעיון של ***התפלגויות-הסתברותיות***. כדי להבין מהי הסתברות, נחשוב על הדוגמא הבאה: נניח שאני רוצה להסתכל על השכר ברוטו בתוך המחלקה הכלכלית של משרד BIG4 גדול.

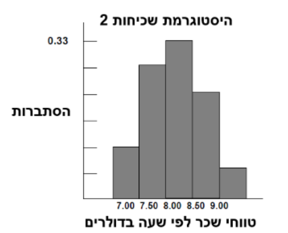
בתור אקטואר ראשית אני אוסף נתונים גולמיים (השכר של כל עובד במחלקה במקרה דנן שלפנינו). שנית, אני מארגן את הנתונים לכדי תצורה בעלת משמעות ובונה תרשים של התפלגות שכיחות עבור הנתונים.

כדי לבנות את ***התפלגות השכיחות***, אני מחלק את השכר למקטעים ומעלה את המקטעים הללו על הציר האופקי של התרשים. לאחר מכן אני רושם את מספר העובדים או יותר נכון את שכיחות העובדים בכל מקטע על הציר האנכי של התרשים.

כעת אני יכול לראות בנקל את התפלגות השכר בתוך המחלקה. מבלי להתעמק בתרשים שלהלן ניתן לראות שרוב העובדים (כ-60 מתוך 180) משתכרים בין 7.00 דולר ל- 9.00 דולר לשעה.



אני יכול להציג את הנתונים הללו גם בתור התפלגות-ההסתברות. התפלגות-הסתברות מציגה את מספר העובדים בכל מקטע כשבר/אחוז ממספר העובדים הכולל. על מנת ליצור התפלגות-הסתברות, אני מחלק את מספר העובדים בכל מקטע במספר העובדים הכולל ורושם את התוצאות על הציר האנכי של התרשים. התרשים הבא מציג את מספר העובדים בכל קבוצת שכר כשבר/אחוז מכלל העובדים. למעשה ניתן לאמוד את הסבירות או את ההסתברות לכך שעובד מסוים אשר יוגרל באקראי מכלל העובדים במחלקה מרוויח שכר במקטע שכר נתון. לדוגמה, בהנחה שאותם התנאים יהיו שרירים וקיימים גם בעת הוצאת המדגם מתוך האוכלוסייה, הרי שקיימת הסתברות של 0.33 (סיכוי של 1 ל- 3) שעובד מסוים אשר יוגרל באקראי מכלל העובדים במחלקה מרוויח בין 8.00 דולר ל-8.50 דולר לשעה.



התפלגויות-הסתברותיות יכולות להיות בדידות או רציפות. התפלגויות-הסתברותיות בדידות מתארות ערכים נפרדים, בדרך כלל מספרים שלמים, ללא ערכי ביניים והן מוצגות כסדרה של מקלות אנכיים. כך למשל, התפלגות-הסתברותית בדידה, עשויה לתאר את מספר הפעמים שהתקבלה התוצאה "פאלי" מתוך ארבע הטלות מטבע הוגן כ-: 0, 1, 2, 3 או 4. התפלגויות-הסתברותיות רציפות הן למעשה הכללות מתמטיות הואיל והן מניחות קיומו של כל ערך ביניים אפשרי בין שני מספרים. כלומר- התפלגות-הסתברותית רציפה מניחה שישנו מספר אינסופי של ערכים בין כל שתי נקודות בהתפלגות. יחד עם זאת, במצבים רבים, ניתן להשתמש ביעילות ב התפלגות-הסתברותית רציפה כדי לבצע קירוב להתפלגות-הסתברותית בדידה למרות שהמודל הרציף לא בהכרח מתאר את המצב במדויק.

לסיכום, סימולציית מונטה קרלו יודעת לייצר ממבתש"ה (משתנים מקריים בלתי-תלויים ושווי-התפלגות) עבור כל הנחת התפלגות. לאמור- לערך האקראי שנבחר עבור ניסוי מסוים אין ולא תהיה שום השפעה על הערך האקראי הבא שנוצר. אני כבר 20 שנה משתמש בסימולציית מונטה קרלו באקסל כאשר אני רוצה לסמלץ תסריטי "What-If" בעולם האמיתי.

הכותב הינו אקטואר, כלכלן אמפירי ומדען נתונים בעל ניסיון רב בפיתוח, יישום ותיקוף מודלים כמותיים בתחומי המימון, הערכות השווי, ניהול הסיכונים, האופציות וההנדסה הפיננסית המסייע לחברות לפתח וליישם מודלים מתקדמים הדורשים הבנה עמוקה בתהליכים סטוכסטיים, ידע בשיטות נומריות ושליטה ברמה גבוהה בכלים כגון R ו- Python ומייעץ בניתוחים כמותיים מתקדמים בתחומים של הנדסה פיננסית, יישום מודל מונטה-קרלו, תהליכים סטוכסטיים ופתרון בעיות כמותיות באמצעות שיטות נומריות מתקדמות.