מבוא לסטטיסטיקה והסתברות / ד"ר דוד סרד

שיעורים 5-6: התפלגויות רציפות, ומשפט הגבול המרכזי

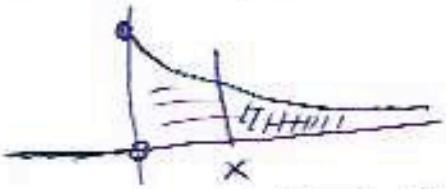
עד עכשיו ראינו התפלגויות בדידות, בכך שסכם כל ההסתברויות של ערכי המשתנה הבדיד היה 1 $\sum p(x)=1$, ועבור כל ערך למשתנה הבדיד היה ערך להסתברות שלו. במקרה של משתנה רציף כמו גובה, גובה, שכר וכו', ההסתברות מתיחסת לתחום שבין 2 ערכים, ונראה אחר שעבור ערך מסוים במשתנה הרציף תמיד ההסתברות היא אפס.

דוגמאות להתפלגויות רציפות, שנטפל בהן יהיו ההתפלגות המעריכית וההתפלגות הנורמלית, שאפשר לייחס להן פונקציה מתימטית רציפה.

התפלגות מעריכית:

$$f(x)=\begin{cases}\frac{1}{k}e^{-\frac{x}{k}} & 0\le x<\infty\\ 0 & \text{אחרת}\end{cases}$$

צורת הפונקציה המעריכית היא, כך שההסתברות בתחום הערכים השליליים היא אפס, ובצד הימני התחום היא יורדת בצורה מעריכית



דרור שהשטח מתחת לעקומה, המכיל את כל הערכים האפשריים צריך להיות 1 (100%). השטח מתחת לעקומה הוא האינטגרל

$$\boxed{\text{אל הוא מחוצה ההתפלגות}}\quad \int_0^\infty \frac{1}{k}e^{-\frac{x}{k}}\,dx=\frac{1}{k}\left[e^{-\frac{x}{k}}\Big/-\frac{1}{k}\right]_0^\infty=-e^{-\frac{x}{k}}\Big]_0^\infty=-(0-1)=1$$

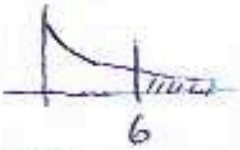
מה שמעניין אותנו הוא השטח מאפס ועד לx ("הגול")
והשטח מ x ועד הסוף (הצנב)

$$\int_0^x \frac{1}{k}e^{-\frac{x}{k}}=\frac{1}{k}\left[e^{-\frac{x}{k}}\Big/-\frac{1}{k}\right]_0^x=-e^{-\frac{x}{k}}\Big]_0^x=-\left[e^{-\frac{x}{k}}-1\right]=1-e^{-\frac{x}{k}}$$

מה שמשרת לצורך חישובים הוא ההסתברות של "הגול" וההסתברות של "הצנב" שהוא במובן המלא $e^{-\frac{x}{k}}$. משתמשים בהתפלגות מעריכית לצורך הערכת אורך חיים של מכשירים אלקטרונים, לדוגמא.

תרגיל: ידוע שאורך חיים ממוצע של טלוויזיה הוא 5 שנים. מה ההסתברות שמכשיר טלוויזיה אקראי "יחיה" מעל 6 שנים, מתחת ל 3 שנים, בין 3 ל 6 שנים.

א. מעל 6 שנים יש למצוא את שטח "הכנף"

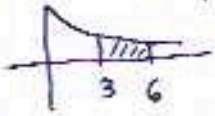


$$e^{-\frac{x}{\lambda}} = e^{-\frac{6}{5}} = e^{-1.2} = 0.3012$$

ב. מתחת ל 3 שנים יש למצוא שטח "הבול"

$$1 - e^{-\frac{3}{5}} = 1 - e^{-0.6} = 0.4512$$

ג. בין 3-6 שנים מבוקש השטח של הכנף מ 3 ואולם פחות השטח של הכנף מ 6 ואולם.



$$e^{-3/5} - e^{-1.2} = 0.5488 - 0.3012 = 0.2476$$

דרור שלפשר לעשות את החישוב גם דרך "הבול" כלומר השטח מ 0 ועד 6 פחות השטח מ 0 ועד 3

$$(1 - e^{-1.2}) - (1 - e^{-3/5}) =$$

נראים שהאחד מינוס, לאחר קיצוץ מקבלים אותה תוצאה, לכן עדיף תמיד לעבוד דרך "הכנף" שאז חוסכים סדרות אפשרות בגלל המינוסים.

נראים, שניתן להשתמש בנוסחאות שטח "הכנף" ושטח הבול ואין צורך לעשות בכל פעם אינטגרל מתאים.

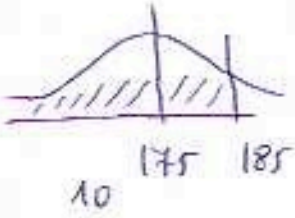
מקודם, אמרנו שההסתברות היא להפרש של 2 ערכי משתנה ולא לערך בודד. ערך בודד יניב הסתברות אפס. למשל מה ההסתברות לאורך חיים של בדיוק 6 שנים. כמו אינטגרל מ 6 עד 6 כלומר אפס. ניתן מעבר להסבר המתימטי הסבר היוני. נניח שנחפש את ההסתברות עבור אורך חיים בין 5.5 ל 6.5 נקבל שטח מסוים. נקטין את התחום לבין 5.9-6.1. ההסתברות תקטן נקטין ל 5.99-6.01 ההסתברות תקטן עוד יותר, ובערך 6 בדיוק תשאל לאפס. לכן במשמעות כבול ההסתברות היא תמוך עבור תחום, ועבור ערך מסוים בדיוק היא תמוך אפס.

בהתפלגות נורמלית: צורתה צורת פעמון

ומאפיינת התפלגות של הרבה תופעות. הוא סימטרי ובכל שמתרחקים מהמרכז (שבו מקום הממוצע = החציון = השכיח), ההסתברות הולכת וקטנה.

הסבלי לחריגים מצד הצדדים הוא תמוך אפשרי (למשל בגובה, הסבלי לגובה במו של צוות מלך הבשן" הוא אפסי) אבל הולך וקטן.

האינטגרל, המתאר את ההתפלגות הנורמלית, הוא קשה לפתרון, ולכן נתנו חשובים לשטח באמצעות טבלה. השמוש בטבלה מתבצע באמצעות **ציון התקן** z $z=\frac{x_i-\bar{x}}{\sigma}$, בלומר מספר סטיות תקן מן המרכז. עבור כל ערך של z ניתן בטבלה שטח (הסתברות) שנרוץ לסמן אותו באות היונית ϕ, והוא השטח מן הקצה השמאלי ועד ל z. למשל: נתון שגובה אוכלוסיה מתפלג נורמלית עם ממוצע 175 וסטיית תקן 10. מצא מה אחוז האוכלוסיה שגובהו מתחת ל 185.



התשובה נעשה תמוך באמצעות ציון התקן = מספר סטיות תקן מהממוצע = z.

$$z=\frac{185-175}{10}=1$$

עבור $z=1$ מסתכלים בטבלה ומקבלים ערך $\phi=0.8413$ שהוא השטח מהקצה השמאלי ועד $z=1$.

בטבלה נותנת ערכי ϕ (שטח) עבור ערכי z בפיצוק של מאות (2 ספרות אחרי הנקודה) וערכי השטח הם בפיצוק של 4 ספרות אחרי הנקודה.

בטבלה המקורית נותנת ערכים רק עבור z חיוביים. עבור z שליליים ניתן לחשב לפי הסימטריה. יש לכם באתר גם טבלה שעושה את החישוב עבור ערכי z שליליים, שהם למעשה

$$\phi(-z)=1-\phi(z)$$

כאלו לציין שיש גם טבלה הנותנת שטחים לא מהקצה השמאלי ועד z, אלא מהמרכז ועד z. עדכים בטבלה כן הם כמובן תמיד הערך של Ø מהטבלה המקורית פחות 0.5 (שהוא השטח מהקצה השמאלי ועד למרכז). נניח שהפתרונות בחלבות התרגילים (הבתיבות בכתבי יד) מתבססת על טבלה כזו. ממליץ לפתור לפי הטבלה המקורית (שטח מן הקצה השמאלי) ולראות שהתשובה הסופית זהה.

נפתור עתה שאלה עם סעיפים שונים המתבססת על טבלת ההתפלגות הנורמלית.

בלרם נעשה כאת, נתמקד בשימוש בטבלה המקורית. הסדר השמאלי עם הכותרת z הוא מספר סטיות התקן מן הממוצע ומצוין מספר סטיות תקן בדיוק של עשירית שעבורן Ø (השטח מן הקצה השמאלי המתאר בכותרת של הטבלה בצורה מקומקלת) נמצא בעזרה בשני עם כותרת 0. אם נרצה דיוק של מאות נצרך ימינה בטבלה עד לספרה המתאימה למשל Ø(z=1.15) נמצא בשורה של 1.1 z תוך תנועה ימינה עד לספרה 5 שבכותרת בעמוד Ø(z=1.15)=0.8749. ניתן לקבל את הערכים גם באקסל באמצעות הפונקציה NORMSDIST(1.15)=.8749. אם נתון השטח ומעוניינים לתרגם אותו ל z, ערבדים הפוך. מחפשים את השטח בגוף הטבלה ומתאימים לו את z. למשל אם נתונה הסברות Ø=0.8749 ומחפשים את z z(Ø=0.8749)=1.15. הפונקציה באקסל שעושה כאת היא NORMSINV(0.8749)=1.15.

תרגיל: נתונה אוכלוסיה מפולגת נורמלית עם ממוצע 175 וסטיית תקן 10. נקח אדם אקראי מאוכלוסיה זו. מה ההסתברות שגובהו: א. מעל 190 ב. בין 175 ל 185

ג. בין 185 ל 195 ד. בין 155–165

ה. מתחת ל 160 ו. מה גובה משקולת הדלת בק ש 99% יעברו תחתיו בלי להתפוצץ.

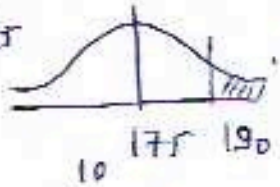
טבלת התפלגות נורמלית סטנדרטית

השטח $\phi(z)$ תחת העקום הנורמלי סטנדרטי משמאל ל- z

z	0	1	2	3	4	5	6	7	8	9
.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8087	.8106	.8133
.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8481	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

לפתרון רצוי לתת שרטוט נפרד לכל סעיף, בכך שניתן לראות בבירור את השטח הנדרש.

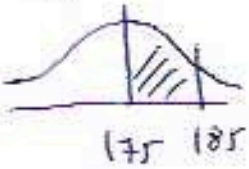
א. $z(190)=\frac{190-175}{10}$ כלומר 1.5 , $\Phi(1.5)=0.9332$



מכאן שהשטח עד 190 הוא 0.9332

והשטח הנדרש מעל 190 יהיה $1-0.9332=0.0668$

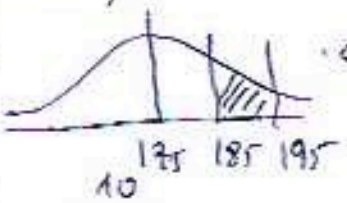
ב. $z(185)=\frac{185-175}{10}=1$, $\Phi(1)=0.8413$, השטח המקומקו



שנדרש יהיה $0.8413-0.5$ שכן עד 175 יש מחצית משטח הפזמון. ברור שאם היינו משתמשים בטבלה הנלווית את השטח מהמרכז ואילך (הטבלה שבה השתמשנו בפתרון השאלה בחלבורת הבתובה בכתב יד) היינו רשמו מקבלים את התוצאה 0.3413 ללא צורך בהחסרה)

ג. $z(195)=\frac{195-175}{10}=2$, $\Phi(2)=0.9772$

$z(185)=1$, $\Phi(1)=0.8413$



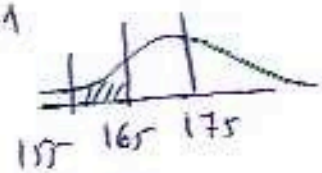
השטח הדרוש הוא ההפרש בין 2 השטחים 0.1359

טעות נפוצה היא שלוקחים $\frac{195-185}{10}=1$ ומתבגמים לשטח. זה במובן לא נכון כי סטיית תקן שקרובה למרכז הרבה ולתר "שמנף" מסטיית תקן הרחוקה מן המרכז.

ד. בין 155-165.

$z(165)=\frac{165-175}{10}=-1$

$z(155)=-2$



ערכים שלילים אינם מופיעים בטבלה המקורית, אלא בטבלה המורחבת הנוספת המופיעה באתר

$\Phi(-1)-\Phi(-2)=0.1587-0.0228=\underline{0.1359}$

אם אין לנו טבלת ערכים שלילים נוכל להשתמש ב $\Phi(-z)=1-\Phi(z)$ כלומר

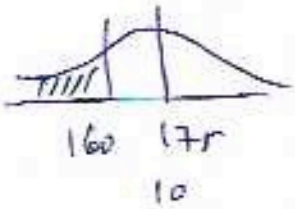
$[1-\Phi(1)]-[1-\Phi(2)]=\Phi(2)-\Phi(1)$

$=0.9772-0.8413=0.1359$

שימו לב שבגלל הסימטריה התוצאות בג' וד' זהות

ה. מתחת ל 160

$z = \frac{160-175}{10} = -1.5$



$\phi(-1.5) = 1 - \phi(1.5)$
$= 1 - 0.9332 = 0.0668$

זה השטח המקווקו.

ו. כאן השאלה הפוכה נתון השטח (ההסתברות) 0.99. יש למצוא את ה z המתאים ולפי "בגם" אותו לגבהה

$\phi(z) = 0.99 \longrightarrow z = 2.33$

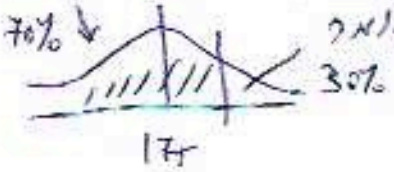
מחפשים בטבלה שטח 0.99 ורואים שזה מתאים ל $z=2.33$.

מכאן $z = 2.33 = \frac{x-175}{10} \Leftarrow x = 198.3$

תרגיל נוסף: באוכלוסיה המתבלטת נורמלית עם גבהה 175 וסטיית תקן 10 אבל מהו העשירון ה-7 ומהו הטווח הבינרבעוני.

א. העשירון השביעי הוא נקודה ש 70% מתחתיה ו 30% מעליה

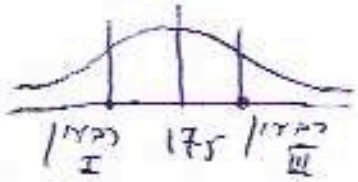
$z = 0.525 \Leftarrow \phi(z) = 0.7$



$z = 0.525 = \frac{x-175}{10}$

$x = 180.25$

ב. מהו הטווח הבינרבעוני



השטח עד לרבעון השלישי $\phi = 0.75$ $z = 0.675$

לכן $x_{75\%} = 181.75$ כי $0.675 = \frac{x-175}{10}$

השטח עד לרבעון הראשון $\phi = 0.25$

$z(\phi = 0.25) = -z(\phi = 0.75) = -0.675$

מכאן $-0.675 = \frac{x-175}{10}$

$x_{25\%} = 168.25$

הטווח הבינרבעוני הוא

רבעון I − רבעון III

$= 181.75 - 168.25$
$= 13.5$

כל דרך נכונה אבל אפשרית ומסוגלת

$z = 0.675$ מתאים לרבעון III

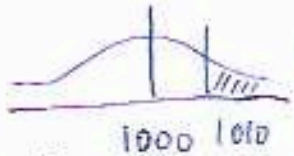
$z = -0.675$ מתאים לרבעון I

כלומר מרחק בין הרבעונים 1.35 סטיות תקן. כאן שבל סטיית תקן היא 10 זה יהיה המרחק 13.5 סמ

משפט הגבול המרכזי:

כאן משפט בסיסי המקשר בין מושגיי המדגם למושגיי האוכלוסיה.
אם יש ידועה אוכלוסיה, אשר לה תוחלת (ממוצע) μ וסטיית תקן σ
(שונות σ^2), והיא מתפלגת נורמלית, אזי אם נקח מאוכלוסיה זו מדגם
בגודל n תהיה תוחלת המדגם μ (כלומר אם נקח את כל המדגמים
האפשריים ונעשה לכל הממוצעים שלהם ממוצע) נקבל μ ושונות $\frac{\sigma^2}{n}$
(סטיית תקן $\frac{\sigma}{\sqrt{n}}$). אנו בעצם מקבלים עבור מדגם מסוים אומדן
נקודתי (המשתנה ממדגם למדגם, אבל התוחלת של הממוצע קבועה,
תמיד μ).
תוחלת סכום ערכי המדגם יהיה $n\mu$, ושונות של סכום כה תהיה $n\sigma^2$
(סטיית תקן $\sigma\sqrt{n}$).
אם לא ידוע שהאוכלוסיה מתפלגת נורמלית אזי אם המדגם מספיק
גדול (מספיק גדול נקבע לפי כלל אצבע, יש קובעים $n \geq 40$
ויש לוקחים $n \geq 35$) תהיה תוחלתו μ וסטיית התקן $\frac{\sigma}{\sqrt{n}}$
כלומר משפט הגבול המרכזי מתקיים עבור מדגמים גדולים.
משפט הגבול המרכזי מאפשר להעריך הסתברויות לממוצע של
מדגם בגודל מסוים מנתוני האוכלוסיה הידועים.
תרגיל: משקל שקית קמח הוא 1000 גרם בממוצע (תוחלת)
עם סטיית תקן 10 גרם. נלקחה שקית אקראית מה
ההסתברות שמשקלה יהיה א. מעל 1010 גרם
ב. בין 1010 ל 1020 גרם ג. מתחת ל 980 גרם

מתוך משפט הגבול המרכזי בהנחת התפלגות
נורמלית יהיה ממוצע המדגם 1000
וסטיית התקן $\frac{\sigma}{\sqrt{n}}$ במקרה שלנו 10



(א)

$$\frac{\sigma}{\sqrt{n}} = \frac{10}{\sqrt{1}} = 10$$

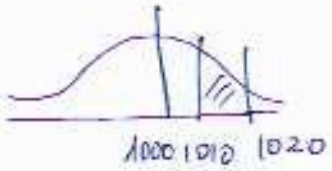
החישוב נעשה באמצעות תפגום ל z (מספר סטיות התקן מהממוצע)

$$z = \frac{1010 - 1000}{10} = 1 \rightarrow \Phi(1) = 0.8413$$ ומשטח הנפרש $1 - 0.8413 = 0.1587$

ד.

$\phi(2) = 0.9772 \leftarrow \quad z(1020) = 2$

$\phi(1) = 0.8413 \qquad z(1010) = 1$



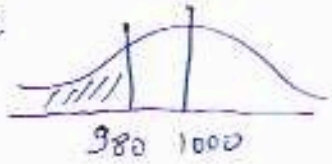
$\underline{\underline{0.1359}}$ התשלום המקווה

$\frac{10}{\sqrt{n}} = 10$

ג. מתחת ל 980 גרם

$\phi(-2) = 1 - \phi(2) =$

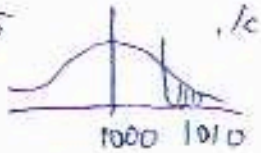
$= 1 - 0.9772 = \underline{\underline{0.0228}} \leftarrow z = \frac{980-1000}{10} = -2$



10

נשאל ריתב אותם שאלות אבל ערכי מדגם בגודל $n = 25$ ולכן ערכי מדגם של שקיות אחת. המקרה כה לפי משפט הגבול המרכזי תהיה חוחלת (ממוצע) המדגם 1000 וסטיית התקן $\frac{\sigma}{\sqrt{n}} = \frac{10}{\sqrt{25}} = 2$

א. $z = \frac{1010-1000}{2} = 5 \qquad \phi(5) = 1$



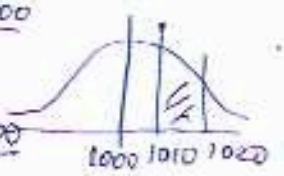
מאחר z גדול רוצים בטבלה שמגדירים במדויק לקבלה הומני של הפרמן, כלומר

$1 - 1 = 0$ הצנה שפוגע משלח הנדרש יהיה

ב

ההסתברות שממוצע מדגם של 25 שקיות יהיה מעל 1010 הוא בקיצור וכה ברור שברור לא יתכן שאם הממוצע הוא 1000 ונקח 25 שקיות יהיה הממוצע שלהם מעל 1010

ד. $z(1020) = \frac{1020-1000}{2} = 10 \qquad \phi(10) = 1$



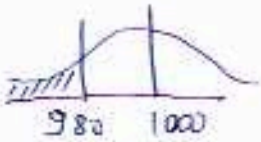
$z(1010) = \frac{1010-1000}{2} = 5 \qquad \phi(5) = 1$

לכן התשלום המקווה הוא 0

ונקח כאן ברור מדוע.

ה. כנ"ל לגבי ג

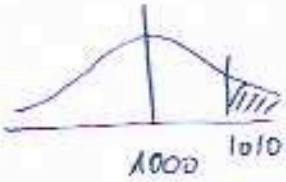
$z(980) = \frac{980-1000}{2} = -10 \qquad \phi(-10) = 1 - \phi(10) = 1 - 1 = 0$



2

תרגיל נוסף: משקל שקית צריך להיות בממוצע 1000 עם סטיית תקן 10
נלקח מידגם של 4 שקיות. מה ההסתברות שממוצע המידגם
יהיה גדול מ 1010 גרם.

מאפיינים המידגם לפי משפט הגבול המרכזי יהיה
ממוצע 1000 וסטיית התקן $\frac{\sigma}{\sqrt{n}} = 5$



$\frac{10}{\sqrt{4}} = 5$

$z(1010) = \frac{1010-1000}{5} = 2$

$\Phi(2) = 0.9772$

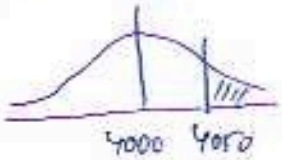
$1 - 0.9772 = 0.0228$ השטח הנגדי (המקומקו)

כלומר סכוי של 2.28% שממוצע המידגם יהיה מעל 1010 צב כמובן
הגיוני לאור כל מה שראינו.

אם ההסתברות שקית אקראית תהיה מעל 1010 היתה 15.87%
ההסתברות שממוצע מידגם בגודל 4 תהיה מעל 1010 היתה 2.28%
ההסתברות שממוצע מידגם בגודל 25 תהיה מעל 1010 היתה אפס.

תרגיל: משקל שקית צריך להיות 1000 עם סטיית תקן 10. מה ההסתברות
שמשקל 4 שקיות יהיה מעל 4050 גרם?

מאפיינים הסכום של המידגם לפי משפט הגבול המרכזי
תוחלת = $n\mu$ = 4000 וסטיית התקן $\sigma\sqrt{n} = \sqrt{4}\cdot 10 = 20$



$z(4050) = \frac{4050-4000}{20} = 2.5$, $\Phi(2.5) = 0.9938$ השטח הנגדי 0.0062

כלומר ההסתברות שמשקלן של 4 שקיות יהיה מעל 4050 הוא 0.62%

תרגיל לפתרון: גל. הראל שבצריכת קוביה בעזרת תבוא התוחלת של
התוצאה המתקבלת 3.5 והשונות 2.9
(רמז $\sum x_i p_i$ = תוחלת, $\sum (x_i - \text{תוחלת})^2 p_i$ = שונות)

ב. נבדק מידגם הכולל הטלת קוביה 100 פעמים.
I. מה ההסתברות שסכום התוצאות בכל 4 הפעמים יהיה בין
342 ל 344. II. גדול מ 345 III. בדיוק 350