1. אם X רציף:

$$\mathbb{E}(Y)=\int_{-\infty}^{\infty}h(x)\,f_X(x)\,dx$$

2. אם X בדיד (מקבל ערכים x_n בהסתברות p_n):

$$\mathbb{E}(Y)=\sum p_n h(x_n)$$

3. אם X מעורב, מפרקים את X למשתנה מקרי רציף X' ומשתנה מקרי בדיד X'':

$$\mathbb{E}(Y)=p\sum p_n h(x_n'')+(1-p)\int_{-\infty}^{\infty}h(x)\,f_X(x)\,dx$$

הוכחה: למקרה הבדיד: $h:\mathbb{R}\to\mathbb{R}$, ומכיוון ש-$h$ איננה חח"ע, יש ל-h אחד מקורות רבים. נסמן y_j את הערכים ש-Y מקבל ונסמן A_j את

$$A_j=\{n:h(x_n)=y_j\}$$

ונסמן ב-q_j את ההסתברות ש-$Y=y_j$.

$$q_j=P(Y=y_j)=P(X\in A_j)=\sum_{i\in A_j}p_i$$

נחשב את התוחלת של Y:

$$\mathbb{E}Y=\sum_j q_j y_j=\sum_j\sum_{n\in A_j}p_n h(x_n)=\sum_n p_n h(x_n)$$

למקרה הרציף: נוכיח למקרה ש-h מונוטונית ממש. נניח ש-h עולה ממש וגזירה.

$$\mathbb{E}Y=\int_{-\infty}^{\infty}y f_Y(y)\,dy=\int_{-\infty}^{\infty}y f_X\left(h^{-1}(y)\right)\left[h^{-1}(y)\right]'dy$$

נציב $y=h(x)$, $dx=\left[h^{-1}(y)\right]'dy$

$$\mathbb{E}Y=\int_{-\infty}^{\infty}h(x)\,f_X(x)\,dx$$

■

הערה 8.4 אם X מעורב ו-$F_X=pF_W+(1-p)F_Z$ כאשר W בדיד ו-Z רציף אז

$$\mathbb{E}X=p\mathbb{E}W+(1-p)\mathbb{E}Z$$

דוגמאות

1. X נורמלי סטנדרטי, $Y = X^2$.

$$\mathbb{E}Y = \int_{-\infty}^{\infty} x^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

2. $Y = |X|$. $h(x) = |x|$.

$$\mathbb{E}Y = \int_{-\infty}^{\infty} |x| \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = 2\int_{0}^{\infty} |x| \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

מטעמי סימטריה ניתן לעשות אינטגרל רק מצד אחד ולהכפיל ב-2

$$\mathbb{E}Y = \left[\frac{2}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\right]_0^{\infty} = \sqrt{\frac{2}{\pi}}$$

3. $Y = e^X$. $h(t) = e^t$

$$\begin{aligned} \mathbb{E}Y &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^x e^{-\frac{x^2}{2}} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\left(x^2-2x+1\right)+\frac{1}{2}} dx \\ &= \frac{\sqrt{e}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(x-1)^2} dx = \sqrt{e} \end{aligned}$$

4. מקרה בדיד: זורקים קוביה הוגנת. אם יוצא $j \in \{1, \dots, 6\}$ משלמים x_j. זו דוגמה מהטיפוס שבו אנחנו מטפלים. נאמר ש-X מקבל את הערכים $\{1, \dots, 6\}$ בהסתברות $\frac{1}{6}$, והפונקציה h היא $h(j) = x_j$. התוחלת של המשתנה $Y = h(X)$ היא התוצאה הצפויה לתשלום.

8.5 תכונות התוחלת

1. אם X משתנה מקרי מנוון ($X \equiv c$) אז $\mathbb{E}X = c$.
2. עבור a קבוע, $Y = aX$ אז $\mathbb{E}Y = a\mathbb{E}X$.
3. עבור a קבוע, $\mathbb{E}(X + a) = \mathbb{E}X + a$
4. אם X משתנה מקרי ונתונות שתי פונקציות g, h אז $\mathbb{E}(g(X) + h(X)) = \mathbb{E}g(X) + \mathbb{E}h(X)$.
הוכחה למקרה ש-X רציף:

$$\begin{aligned} \mathbb{E}(g(X) + h(X)) &= \int_{-\infty}^{\infty} (g(x) + h(x)) f_X(x)\, dx = \\ &= \int_{-\infty}^{\infty} g(x) f_X(x)\, dx + \int_{-\infty}^{\infty} h(x) f_X(x)\, dx \\ &= \mathbb{E}g(X) + \mathbb{E}h(X) \end{aligned}$$

5. לכל X,Y: $\mathbb{E}\left(X+Y\right)=\mathbb{E}X+\mathbb{E}Y$. (4 הוא מקרה פרטי של תכונה זו) (בתנאי של-X ו-Y יש בכלל תוחלת).

6. $X\geq0\Leftarrow\mathbb{E}X\geq0$

7. מסקנה מ-5 ו-6 : אם $X\geq Y$ אז גם $\mathbb{E}X\geq\mathbb{E}Y$

מסקנה 8.5 התוחלת היא אופרציה ליניארית

8.6 המומנט של משתנה מקרי

הגדרה 8.6 יהי X מ"מ ו-$k\geq1$ שלם, אז המומנט ה-k של X הוא

$$m_{k}=M_{k}=\mathbb{E}X^{k}$$

דוגמאות

1. נניח $X\sim\text{Pois}\left(\lambda\right)$ ונחשב את M_{2}.

$$\begin{aligned}
\mathbb{E}X^{2} & = \sum x_{n}^{2}p_{n}=\sum_{n=0}^{\infty}n^{2}e^{-\lambda}\frac{\lambda^{n}}{n!}=e^{-\lambda}\sum_{n=0}^{\infty}\left(n\left(n-1\right)\frac{\lambda^{n}}{n!}+n\frac{\lambda^{n}}{n!}\right)\\
 & = e^{-\lambda}\sum_{n=0}^{\infty}n\left(n-1\right)\frac{\lambda^{n}}{n!}+e^{-\lambda}\sum_{n=0}^{\infty}n\frac{\lambda^{n}}{n!}=e^{-\lambda}\lambda^{2}\left(\sum_{n=0}^{\infty}\frac{\lambda^{n}}{n!}\right)''+e^{-\lambda}\lambda\left(\sum_{n=0}^{\infty}\frac{\lambda^{n}}{n!}\right)'\\
 & = \lambda^{2}+\lambda
\end{aligned}$$

2. $X\sim N\left(0,1\right)$. נחשב את M_{k}:

$$\begin{aligned}
M_{k} & = \mathbb{E}X^{k}=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}x^{k}e^{-\frac{x^{2}}{2}}dx\\
 & = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}x^{k-1}xe^{-\frac{x^{2}}{2}}dx=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}x^{k-1}\left(-e^{-\frac{x^{2}}{2}}\right)'dx\\
 & = \frac{1}{\sqrt{2\pi}}\left[\left(-x^{k-1}e^{-\frac{x^{2}}{2}}\right)_{-\infty}^{\infty}+\int_{-\infty}^{\infty}\left(k-1\right)x^{k-2}e^{-\frac{x^{2}}{2}}dx\right]\\
 & = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\left(k-1\right)x^{k-2}e^{-\frac{x^{2}}{2}}dx=\frac{\left(k-1\right)}{\sqrt{2\pi}}\int_{-\infty}^{\infty}x^{k-2}e^{-\frac{x^{2}}{2}}dx\\
 & = \left(k-1\right)M_{k-2}=\left(k-1\right)\mathbb{E}X^{k-2}
\end{aligned}$$

אם $k=2l$ זוגי אז

$$\mathbb{E}X^{k}=\left(2l-1\right)\left(2l-3\right)\cdots1=\left(2l-1\right)!!$$

ואם k אי זוגי המומנט הוא אפס.

8.7 שונות

$X\sim N\left(\mu,\sigma^2\right)$. אנחנו מחפשים מדד מספרי למידת הפיזור של X סביב התוחלת שלו. $\mathbb{E}\left(X-\mu\right)=0$ ולכן לא יתאים למטרה. יותר הגיוני למדוד $\mathbb{E}\left|X-\mu\right|$ אבל קשה לחשב אותו. מסיבות טכניות ואחרות מסתכלים על

$$\mathbb{E}\left(X-\mu\right)^2$$

נוסחה זו נותנת משקל גדול יותר לשגיאות גדולות מאשר לשגיאות קטנות.

הגדרה 8.7 יהי X מ"מ עם תוחלת $\mathbb{E}X=\mu$. אז השונות (**Variance**) של X היא

$$\mathrm{Var}\left(x\right)=\sigma_X^2=\sigma^2=\mathbb{E}\left(X-\mu\right)^2$$

סטיית התקן מסומנת σ.

8.7.1 תכונות השונות

נחשב את השונות סביב נקודה כלשהי, לאו דווקא התוחלת של X:

$$\begin{aligned}\mathbb{E}\left(X-a\right)^2 &= \mathbb{E}\left(\left(X-\mu\right)+\left(\mu-a\right)\right)^2=\\ &= \mathbb{E}\left(X-\mu\right)^2+\underbrace{2\mathbb{E}\left(\mu-a\right)\left(X-\mu\right)}_{0}+\mathbb{E}\left(\mu-a\right)^2\\ &= \mathbb{E}\left(X-\mu\right)^2+\left(\mu-a\right)^2\end{aligned}$$

מסקנות:

1. $\mathbb{E}\left(X-a\right)^2\geq\mathbb{E}\left(X-\mu\right)^2$, ושוויון מתקיים רק אסם $a=\mu$. יוצא שאם נמדוד את הסטיות מכל נקודה a, אז הסטיה המינימלית תתקבל ב-$a=\mu$.
2. נציב $a=0$:

$$\mathbb{E}X^2=\mathbb{E}\left(X-\mu\right)^2+\mu^2=\mathrm{Var}\left(X\right)+\left(\mathbb{E}X\right)^2$$

כלומר

$$\mathrm{Var}\left(X\right)=\mathbb{E}X^2-\left(\mathbb{E}X\right)^2$$

השונות שווה למומנט השני פחות התוחלת בריבוע.

3. השונות היא תוחלת של משתנה אי שלילי, ולכן היא אי שלילית. $\mathrm{Var}\left(X\right)\geq 0$

$$\begin{aligned}\mathbb{E}X^2-\left(\mathbb{E}X\right)^2 &\geq 0\\ \mathbb{E}X^2 &\geq \left(\mathbb{E}X\right)^2\end{aligned}$$

לכל X עבורו קיימים מומנט שני ותוחלת. למעשה אם נחליף את X ב-$|X|$ אז נקבל

$$\mathbb{E}\left|X\right|\leq\sqrt{\mathbb{E}X^2}$$

ואם ל-X יש מומנט שני אז יש לו גם תוחלת (התוחלת של X מתכנסת בהחלט).

4. נניח $X \equiv c$ משתנה מקרי מנוון, $\mathrm{Var}\left(X\right)=0$.

5. לכל a קבוע,

$$\mathrm{Var}\left(X+a\right)=\mathrm{Var}\left(X\right)$$

כי

$$\mathrm{Var}\left(X+a\right)=\mathbb{E}\left(\left(X+a\right)-\mathbb{E}\left(X+a\right)\right)=\mathbb{E}\left(X+a-a-\mathbb{E}X\right)=\mathbb{E}\left(X-\mathbb{E}X\right)=\mathrm{Var}\left(X\right)$$

6. אם a קבוע אז

$$\mathrm{Var}\left(aX\right)=a^{2}\mathrm{Var}\left(X\right)$$

$$\sigma_{aX}=\left|a\right|\sigma_{X}$$

דוגמאות

1. $X\sim\mathrm{Pois}\left(\lambda\right)$.

$$\mathrm{Var}\left(X\right)=\mathbb{E}X^{2}-\left(\mathbb{E}X\right)^{2}=\lambda^{2}+\lambda-\lambda^{2}=\lambda$$

2. $X\sim\mathrm{Bin}\left(n,p\right)$

$$\mathbb{E}X=np$$

נחשב את המומנט השני של X:

$$\begin{aligned}\mathbb{E}X^{2} &= \sum_{j=0}^{n}j^{2}\binom{n}{j}p^{j}q^{n-j}=p^{2}\sum_{j=0}^{n}j\left(j-1\right)\binom{n}{j}p^{j-2}q^{n-j}+p\sum_{j=0}^{n}j\binom{n}{j}p^{j-1}q^{n-j}\\ &= p^{2}\left(\sum_{j=0}^{n}\binom{n}{j}p^{j}q^{n-j}\right)''+p\left(\sum_{j=0}^{n}\binom{n}{j}p^{j}q^{n-j}\right)'\\ &= p^{2}\left[\left(p+q\right)^{n}\right]''+p\left[\left(p+q\right)^{n}\right]'=p^{2}n\left(n-1\right)+pn\end{aligned}$$

ולכן השונות של X היא

$$\begin{aligned}\mathrm{Var}\left(X\right) &= \mathbb{E}X^{2}-\left(\mathbb{E}X\right)^{2}=p^{2}n\left(n-1\right)+pn-\left(pn\right)^{2}\\ &= p^{2}n^{2}-p^{2}n+pn-p^{2}n^{2}=pn\left(1-p\right)=npq\end{aligned}$$

3. $X = \text{Geom}(p)$, $\mathbb{E}X = \frac{1}{p}$.

$$\begin{aligned}
\mathbb{E}X(X-1) &= \sum_k k(k-1)pq^{k-1} \\
&= pq\sum_k k(k-1)pq^{k-2} = pq\left(\sum_k q^k\right)'' \\
&= pq\left(\frac{1}{1-q}\right)'' = pq\left(\frac{2}{(1-q)^3}\right) = \frac{2pq}{p^3} = \frac{2q}{p^2}
\end{aligned}$$

לכן

$$\begin{aligned}
\text{Var}(X) &= \mathbb{E}X^2 - (\mathbb{E}X)^2 = \mathbb{E}X(X-1) + \mathbb{E}X - (\mathbb{E}X)^2 \\
&= \frac{2q}{p^2} + \frac{1}{p} - \frac{1}{p^2} = \frac{2q+p-1}{p^2} = \frac{q}{p^2}
\end{aligned}$$

4. $X \sim U(0,1)$,

$$\mathbb{E}X^2 = \int_0^1 x^2 \cdot 1dx = \frac{1}{3}$$

$$\text{Var}(X) = \frac{1}{3} - \left(\frac{1}{2}\right)^2 = \frac{1}{12}$$

אם $X \sim U(a,b)$ אז

$$\text{Var}(X) = \text{Var}(X-a) = (b-a)^2\,\text{Var}\left(\frac{X-a}{b-a}\right) = \frac{(b-a)^2}{12}$$

5. $X \sim \exp(\lambda)$, $\mathbb{E}X = \frac{1}{\lambda}$

$$\mathbb{E}X^2 = \int_0^\infty x^2\lambda e^{-\lambda x}dx = \frac{2}{\lambda^2}$$

$$\text{Var}(X) = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$$

6. $X \sim N(0,1)$, $\mathbb{E}X = 0$, $\mathbb{E}X^2 = 1$

$$\text{Var}(X) = 1 - 0^2 = 1$$

אם $X \sim N(\mu,\sigma^2)$:

$$\frac{X-\mu}{\sigma} \sim N(0,1)$$

או אם $Z \sim N(0,1)$ אז

$$X = \sigma Z + \mu$$

לכן

$$\begin{aligned}\mathbb{E}X &= \mu \\ \mathrm{Var}(X) &= \sigma^2 \mathrm{Var}(Z) = \sigma^2\end{aligned}$$

ניתן היה גם לחשב את השונות ישירות:

$$\mathrm{Var}(X) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} (x-\mu)^2 e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

8.7.2 הקשר בין המומנט השני לטורי פוריה

נניח ש-$X \sim U(o, 2\pi)$, $f_X(x) = \frac{1}{2\pi}$ ($0 \le x \le 2\pi$), ונתונה $h(t)$ אז,

$$\begin{aligned}\mathbb{E}h(X) &= \frac{1}{2\pi}\int_0^{2\pi} h(x)\,dx \\ \mathbb{E}X^2 &= \frac{1}{2\pi}\int_0^{2\pi} x^2 dx \\ \mathbb{E}h^2(X) &= \frac{1}{2\pi}\int_0^{2\pi} h^2(x)\,dx = ||h||^2_{L^2(0,2\pi)}\end{aligned}$$

באופן כללי אם X משתני מקרי עם צפיפות אז

$$\mathbb{E}h^2(X) = \int_{-\infty}^{\infty} h^2(x)\,dx$$

8.8 אי שיוויון צ׳בישב

אינטואיטיבית, אם נבחר X רחוק מהממוצע, סביר שההסתברות לכך תהיה קטנה יותר מאשר X קרוב לממוצע.

משפט 8.8 יהי X מ"מ עם תוחלת μ ושונות σ^2, אז לכל מספר חיובי מתקיים

$$P(|X-\mu| > b) < \frac{\sigma^2}{b^2}$$

מכונה אמורה ליצר מוטות באורך מטר. בפועל המוטות הם בעלי אורך מקרי X. $\mathbb{E}X =$ 1. ידוע לנו גם כי $\sigma_X = \frac{1}{1000}$.

מהי ההסתברות שהסטיה גדולה מסנטימטר?

$$P\left(|X-1| > \frac{1}{100}\right)$$

אי שיוויון נותן הערכה בלבד:

$$P\left(|X-1|>\frac{1}{100}\right)<\frac{\left(\frac{1}{1000}\right)^2}{\left(\frac{1}{100}\right)^2}=\frac{1}{100}$$

הוכחה: נניח ש-X מ"מ רציף עם צפיפות $f_X\left(x\right)$.

$$\begin{aligned}P\left(|X-\mu|>b\right) &= \int_{\{|X-\mu|>b\}} f_X\left(x\right)dx \\ &\leq \int_{\{|X-\mu|>b\}} \frac{\left(x-\mu\right)^2}{b^2} f_X\left(x\right)dx\end{aligned}$$

מכיוון ש-$\frac{x-\mu}{b}$ גדול מאחד בתחום.
נגדיל את התחום לכל הישר, וכך נגדיל עוד יותר את האינטגרל:

$$P\left(|X-\mu|>b\right)\leq\frac{1}{b^2}\int_{-\infty}^{\infty}\left(x-\mu\right)^2 f_X\left(x\right)dx$$

עבור $h\left(x\right)=\left(x-\mathbb{E}X\right)^2$ קיבלנו בעצם את התוחלת של המשתנה המקרי $\mathbb{E}h\left(X\right)$:

$$P\left(|X-\mu|>b\right)\leq\frac{1}{b^2}\mathbb{E}\left(X-\mathbb{E}X\right)^2=\frac{\sigma^2}{b^2}$$

■

במשתנים מקריים ספציפיים אפשר לקבל קצב שאיפה לאפס (כאשר $b\to\infty$) מהיר בהרבה.
נניח ש-$X\sim N\left(0,1\right)$:

$$\begin{aligned}P\left(|X-\mu|>b\right) &= P\left(|X|>b\right) \\ &= \frac{2}{\sqrt{2\pi}}\int_0^{\infty}e^{-\frac{x^2}{2}}dx<\frac{2}{\sqrt{2\pi}}\int_0^{\infty}\frac{e^x}{e^b}e^{-\frac{x^2}{2}}dx\end{aligned}$$

כי $\frac{x}{b}>1$.

$$P\left(|X|>b\right)\leq e^{-b}\underbrace{\frac{2}{\sqrt{2\pi}}\int_0^{\infty}e^{x-\frac{x^2}{2}}dx}_{k}=\frac{k}{e^b}$$

לכן קצב הירידה הוא אקספוננציאלי ולא ריבועי.

8.9 אי שיוויון ינסן

ידוע לנו ש-

$$\mathbb{E}X^2\geq\left(\mathbb{E}X\right)^2$$

כי ההפרש בינהם הוא השונות והיא מספר חיובי.
עבור $h(t)=t^2$ ניתן לרשום את אי השיוויון בתור:

$$\mathbb{E}h(X)\geq h(\mathbb{E}X)$$

משפט 8.9 אם h פונקציה קמורה אז לכל משתנה מקרי X מתקיים $\mathbb{E}h(X)\geq h(\mathbb{E}X)$.
לדוגמה, $h(t)=|t|^n$ קמורה לכל $n>1$ לכן

$$\mathbb{E}X^n\geq(\mathbb{E}X)^n$$

גם $|t|^\alpha$ קמורה עבור $\alpha>1$.

$$(\mathbb{E}|X|^n)^{\frac{1}{n}}\geq(\mathbb{E}|X|^m)^{\frac{1}{m}}\geq(|X|^m)^{\frac{1}{n}}$$

הגדרה 8.10 פונקציה $f:\mathbb{R}\to\mathbb{R}$ היא קמורה אם היא פונקציה שבה לכל שתי נקודות a,b הגרף של f בקטע $[a,b]$ נמצא מתחת למיתר המקשר את $(a,f(a))$ עם $(b,f(a))$.
לדוגמה, עבור נקודה האמצע של המיתר:

$$f\left(\frac{a+b}{2}\right)\leq\frac{1}{2}f(a)+\frac{1}{2}f(b)$$

באופן כללי לכל $\lambda,\mu\geq0$ כך ש-$\lambda+\mu=1$

$$f(\lambda a+\mu b)\leq\lambda f(a)+\mu f(b)$$

אם f גזירה פעמים אז f קמורה $\Leftrightarrow f''\geq0$.

הוכחה: הוכחת אי שיוויון ינסן:
נכתוב את הגדרת הפונקציה הקמורה בסימון שונה. נניח שהנקודות הן a_1,a_2 והמשקלות הן λ_1,λ_2 כלומר $\lambda_i\geq0$ ו-$\lambda_1+\lambda_2=1$. אז

$$h(\lambda_1a_1+\lambda_2a_2)\leq\lambda_1h(a_1)+\lambda_2h(a_2)$$

באינדוקציה מוכיחים שאם h קמורה אז לכל n ולכל $a_1,\ldots,a_n$ ולכל $\lambda_1,\ldots,\lambda_n$ כאשר $\lambda_1,\ldots,\lambda_n\geq0$ ו-$\sum_{i=1}^n\lambda_i=1$ מתקיים

$$h\left(\sum_{i=1}^n\lambda_ia_i\right)\leq\sum_{i=1}^n\lambda_ih(a_i)$$

אם X דיסקרטי עם מספר סופי של ערכים, נניח ש-X מקבל את הערכים $x_1, \dots x_n$ בהסתברות $p_1, \dots p_n$ בהתאמה. אז

$$\begin{aligned} \mathbb{E}X &= \sum_{i=1}^{n} p_i x_i \\ \mathbb{E}h\left(X\right) &= \sum_{i=1}^{n} p_i h\left(x_i\right) \end{aligned}$$

והטענה נובעת מהגדרת הקמירות.

$$h\left(\mathbb{E}X\right) = h\left(\sum_{i=1}^{n} p_i x_i\right) \leq \sum_{i=1}^{n} p_i h\left(x_i\right) = \mathbb{E}h\left(X\right)$$

■

8.10 פונקציה יוצרת מומנטים

נתון מ"מ X, אז הפונקציה יוצרת המומנטים שלו היא

$$M_X\left(s\right) = \mathbb{E}e^{sX}$$

היא מוגדרת עבור s ממשי בתנאי של-e^{sX} יש תוחלת.

דוגמאות $X \sim \exp\left(\lambda\right)$

$$\begin{aligned} M_X\left(s\right) &= \mathbb{E}e^{sX} \\ &= \int_0^{\infty} e^{sx} \lambda e^{-\lambda x} dx = \lambda \int_0^{\infty} e^{(s-\lambda)x} dx \\ &= \frac{\lambda}{s-\lambda}\left[e^{(s-\lambda)x}\right]_0^{\infty} = \begin{cases} \frac{\lambda}{s-\lambda} & s < \lambda \\ \infty & s \geq \lambda \end{cases} \end{aligned}$$

$X \sim \mathrm{Geom}\left(p\right)$

$$\begin{aligned} M_X\left(s\right) &= \sum_{n\geq 1} e^{sn} p q^{n-1} \\ &= p e^s \sum_{n \geq 1} e^{s(n-1)} q^{n-1} \\ &= p e^s \sum_{n\geq 1} \left(e^s q\right)^{n-1} = \begin{cases} \frac{pe^s}{1-e^s q} & s < -\log q \\ \infty & e^s q \geq 1 \end{cases} \end{aligned}$$

מהי הצורה הכללית של M_X כאשר יש ל-X צפיפות?

$$M_X\left(s\right) = \mathbb{E}e^{sX} = \int_{-\infty}^{\infty} e^{sx} f_X\left(x\right) dx$$

תוצאה זו מאוד דומה לטרנספורם לפלס של f_X.

8.10.1 תחום ההגדרה של M_X

$M_X(0)=1$ לכל X.

טענה 8.11 תחום ההגדרה של M_X הוא תמיד קטע (כולל אפשרות של קרן אינסופית או כל הישר).

כלומר, תחום ההגדרה הוא קבוצה קמורה בישר.

הוכחה: נקבע שתי נקודות s_1,s_2 עבורם $M_X(s_i)<\infty$.

נבחר $s_1<s<s_2$ ונציג אותו כממוצע של s_1,s_2:

$$s=\lambda s_1+(1-\lambda)s_2$$

כאשר $0<\lambda<1$.

נשתמש בקמירות של הפונקציה $h(t)=e^t$.

$$\begin{aligned} M_X(s) &= \mathbb{E}e^{sX}= \\ &= \mathbb{E}e^{\lambda s_1 X+(1-\lambda)s_2 X}\le \lambda\mathbb{E}e^{s_1X}+(1-\lambda)\mathbb{E}e^{s_2X} \\ &= \lambda M_X(s_1)+(1-\lambda)M_X(s_2)<\infty \end{aligned}$$

■

נחשב את הנגזרת של $M_X(s)$:

אם X רציף:

$$M_X'(s)=\frac{d}{ds}\int_{-\infty}^{\infty}e^{sx}f_X(x)\,dx=\int_{-\infty}^{\infty}xe^{sx}f_X(x)\,dx=\mathbb{E}Xe^{sX}=\mathbb{E}\left(e^{sX}\right)'$$

אם X בדיד:

$$M_X'(s)=\frac{d}{ds}\sum e^{sx_n}p_n=\sum x_ne^{sx_n}p_n=\mathbb{E}Xe^{sX}=\mathbb{E}\left(e^{sX}\right)'$$

נניח שמותר לגזור בתוך סימן התוחלת, ואז

$$M_X'(s)=\mathbb{E}Xe^{sX}$$

נציב $s=0$ ונקבל

$$M_X'(0)=\mathbb{E}X$$

$$\begin{aligned} M_X''(s) &= \mathbb{E}X^2e^{sX} \\ M_X^{(k)}(s) &= \mathbb{E}X^ke^{sX} \end{aligned}$$

ושוב נציב $s=0$ ונקבל

$$M_X^{(k)}(0)=\mathbb{E}X^k$$

כלומר קיבלנו את המומנט ה-k של X.

משפט 8.12 נניח ש-$M_X(s)$ מוגדרת וגזירה k פעמים בסביבת הנקודה $s=0$, אז יש ל-X מומנט k-י והוא ניתן ע"י:

$$M_X^{(k)}(0) = \mathbb{E}X^k = m_k$$

דוגמה $X \sim \exp(\lambda)$. מהו m_{39}?
כבר חישבנו ש-

$$\begin{aligned} M_X(s) &= \frac{\lambda}{s-\lambda} = \frac{1}{\frac{s}{\lambda}-1} = -\frac{1}{1-\frac{s}{\lambda}} = -\sum_{n=0}^{\infty}\left(\frac{s}{\lambda}\right)^n = -\sum_{n=0}^{\infty}\lambda^{-n}s^n \\ &= \sum_{n=0}^{\infty}\frac{M_X^{(n)}(0)}{n!}s^n \end{aligned}$$

לכן מהשוואת מקדמים נקבל

$$M_X^{(n)}(0) = -\frac{n!}{\lambda^n}$$

8.11 הפונקציה האופיינית (Characteristic Function)

לכל s ממשי נגדיר פונקציה אופיינית

$$\phi_X(s) = \mathbb{E}e^{isX}$$

במקרה הרציף נקבל ש-

$$\phi_X(s) = \int_{-\infty}^{\infty} e^{isx} f_X(x)\,dx$$

וזה דומה לטרנספורם פורייה של פונקצית הצפיפות.

8.11.1 חשיבות הפונקציה האופיינית

1. היא מוגדרת לכל X ולכל s ממשי, כי $\left|e^{isX}\right| = 1$ ולכן גם התוחלת היא מספר שערכו המוחלט הוא לכל היותר אחת.
2. אם לשני משתנים מקריים X ו-Y מתקיים $\phi_X(s) = \phi_Y(s)$ אז גם $F_X = F_Y$.

הערה 8.13 אם ϕ_X גזירה ניתן לגזור בתוך סימן התוחלת

$$\phi_X^{(k)}(s) = i^k\mathbb{E}X^k e^{isX}$$

ולכן

$$\phi_X^{(k)}(0) = i^k\mathbb{E}X^k = i^k m_k$$

9 וקטורים אקראיים

זורקים 2 מטבעות הוגנות ושואלים מהן התוצאות. חייבים לתת אינפורמציה על ההסתב-רויות של כל התוצאות האפשריות, כלומר 4 מספרים $(p_{ij})_{(i,j=0,1)}$שהם ההסתברות שבשלב הראשון תתקבל התוצאה i ובשני j.

$$\begin{pmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{pmatrix}$$

ההסתברות שהמטבע הראשון יהיה 0 הוא סכום ההסתברויות בשורה הראשונה, כלומר $p_{00} + p_{01} = \frac{1}{2}$.

ההסתברות שהמטבע השני יהיה 1 הוא סכום העמודה השניה, וגם הוא $\frac{1}{2}$. למעשה סכום כל עמודה וכל שורה הוא $\frac{1}{2}$. נסמן $p_{00} = t$, אז ניתן לסמן

$$\begin{pmatrix} t & \frac{1}{2} - t \\ \frac{1}{2} - t & t \end{pmatrix}$$

כל זה נובע מכך שהמטבע הוא הוגן. לכל $0 \leq t \leq \frac{1}{2}$ יש ניסוי אפשרי שמתאים לו, כאשר t קובע את הקשר בין המטבעות.

נכליל למקרה כללי יותר. נניח שיש לנו שני משתנים מקריים X ו-Y שמקבלים ערכים $x_1, x_2, \ldots$ ו-$y_1, y_2, \ldots$ בהסתברויות $p_1, p_2, \ldots$ ו-$q_1, q_2, \ldots$ בהתאמה.

נסמן

$$p_{ij} = P\left(X = x_i \wedge Y = y_j\right)$$

$$P\left(X = x_i\right) = \sum_{j=1}^{\infty} P\left(X = x_i \wedge Y = y_j\right) = \sum_{j=1}^{\infty} p_{ij}$$

9.1 וקטור מקרי בדיד

הגדרה 9.1 נאמר שהוקטור המקרי (X, Y) הוא וקטור מקרי בדיד, אם הוא יכול לקבל מספר סופי או בן מניה של ערכים. אם הערכים הם (x_i, y_j), אז לפונקציה $P_{(X,Y)}\left(i, j\right) = P\left(X = x_i \wedge Y = y_j\right)$ קוראים פונקצית ההסתברות של הוקטור המקרי.

בסימונים הנ"ל לפונקצית ההסתברות של X שנוסחתה היא

$$P\left(X = x_i\right) = \sum_j P_{(X,Y)}\left(i, j\right)$$

ולפונקצית ההסתברות של Y שניתנת ע"י

$$P\left(Y = y_j\right) = \sum_i P_{(X,Y)}\left(i, j\right)$$

אנו קוראים פונקציות ההסתברות השוליות.

פונקצית ההתפלגות של הוקטור (X, Y) היא הפונקציה

$$F_{XY}\left(x, y\right) = P\left(X \leq x \ \wedge \ Y \leq y\right)$$

אם $X=(X_1,\ldots,X_n)$ וקטור n-ממדי, אז פונקצית ההסתברות שלו היא

$$P_{X_1,\ldots,X_n}(x_1,\ldots,x_n)=P(X_1=x_1\wedge\cdots\wedge X_n=x_n)$$

ופונקצית ההתפלגות היא

$$F_{X_1,\ldots,X_n}(x_1,\ldots,x_n)=P(X_1\leq x_1\wedge\cdots\wedge X_n\leq x_n)$$

פונקציות ההסתברות השוליות הן

$$P_{X_i}(x)=\sum P_{X_1,\ldots\hat{X}_i,\ldots,X_n}(x_1,\ldots,x,\ldots x_n)$$

כלומר סוכמים את כל האפשרויות בהן המשתנה X_i מקבל את הערך x.

דוגמא (X,Y) וקטור מקרי דיסקרטי, כאשר $P((x,y)=(i,j))=c2^{-j}$ ($1\leq i\leq j<\infty$).

ההסתברות השולית של X היא

$$P_X(i)=\sum_{j\geq i}c2^{-j}=\frac{c}{2^{i-1}}$$

וההסתברוקת השולית של Y:

$$P_Y(j)=\sum_{i=1}^{j}c2^{-j}=jce2^{-j}$$

ניתן לחשב את c על ידי המשוואה $\sum_{i,j}c2^{-j}=1$.
בדרך אחרת:

$$P_X(i)=2c\frac{1}{2}2^{-(i-1)}$$

אבל זוהי בדיוק ההסתברות של משתנה מקרי גיאומטרי עם פרמטר $\frac{1}{2}$. לכן $c=\frac{1}{2}\Leftarrow 2c=1$.

9.2 וקטור אקראי רציף

הגדרה 9.2 וקטור אקראי הוא רציף אם יש פונקציה $f_{X,Y}$ הנקראת פונקצית הצפיפות של הוקטור, כך שלכל $a_1<b_1$ ו-$a_2<b_2$

$$P(a_1\leq X\leq b_1\wedge a_2\leq Y\leq b_2)=\int_{a_1}^{b_1}\int_{a_2}^{b_2}f_{X,Y}(x,y)\,dydx$$

f חייבת להיות חיובית, ו-$\int_{-\infty}^{\infty}\int_{-\infty}^{-\infty}f=1$. f יחידה עד כדי שינוי שלא משנה ערך של אינטגרלים.

הגדרה 9.3 פונקצית ההתפלגות של וקטור רציף היא

$$\begin{aligned} F_{X_1,\ldots,X_n}(x_1,\ldots,x_n) &= P(X_1 \le x_1 \wedge \cdots \wedge X_n \le x_n) \\ &= \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f_{\vec{X}}(t_1,\ldots,t_n)\, dt_n \cdots dt_1 \end{aligned}$$

באופן כללי יותר

$$P\left(\vec{X} \in A\right) = \int_A f_{\vec{X}} dx_1 \cdots dx_n$$

ההתפלגות השולית במקרה הדו ממדי תהיה

$$F_X(x) = \int_{-\infty}^{x} \int_{-\infty}^{\infty} f_{XY}(s,t)\, dt ds$$

נקבל את הצפיפות השולית על ידי גזירה לפי x:

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,t)\, dt$$

9.2.1 תכונות פונקצית ההתפלגות

$$F_{XY}(x,y) = P(X \le x \wedge Y \le y)$$

1. $0 \le F_{XY}(x,y) \le 1$
2. $\lim_{x\to-\infty} F_{XY}(x,y) = 0$ לכל y.
3. $\lim_{x\to\infty} F_{XY}(x,y) = P(Y \le y) = F_Y(y)$
4. עבור הריבוע $(x,y),(x+a,y),(x,y+b),(x+a,y+b)$ ($a,b \ge 0$) ההסתברות שנקבל וקטור בתוך הריבוע היא בוודאי אי שלילית.
הסתברות זו שווה ל-

$$F_{XY}(x+a,y+b) - F_{XY}(x,y+b) - F_{XY}(x+a,y) + F_{XY}(x,y) \ge 0$$

9.3 משתנים מקריים בלתי תלויים

הגדרה 9.4 יהי $(X_1,\ldots,X_n)$ וקטור מקרי עם פונקצית התפלגות משותפת $F_{X_1,\ldots,X_n}$. נאמר שהמ"מ $X_1,\ldots,X_n$ (רכיבי הוקטור) הם בלתי תלויים אם

$$F_{X_1,\ldots,X_n}(x_1,\ldots,x_n)=F_{X_1}(x_1)F_{X_2}(x_2)\cdots F_{X_n}(x_n)$$

למשל בשני ממדים:

$$F_{XY}(x,y)=F_X(x)F_Y(y)$$

כלומר ההסתברות ש-$X\le x$ וגם $Y\le y$ שווה למכפלת ההסתברויות הנ"ל.

$$F_{XY}(x,y)=P(X\le x\wedge Y\le y)=P(X\le x)P(Y\le y)=F_X(x)F_Y(y)$$

ההגדרה למעלה אומרת שמידע על אחד המאורעות $X\le x$ או $Y\le y$ אינו תורם אינפורמציה להסתברות של המאורע האחר. באופן אינטואיטיבי אי התלות פירושה שלכל שתי קבוצות A,B מתקיים

$$P(X\in A\wedge Y\in B)=P(X\in A)P(Y\in B)$$

ולא רק עבור $A=\{X\le x\}$, $B=\{Y\le y\}$.

טענה 9.5 אם (X,Y) ב"ת, הדרישה הזו אכן מתקיימת.

הערה 9.6 אם $X_1,\ldots,X_n$ בלתי תלויים ויש וקטור צפיפות, אז מתקיים

$$f_{X_1,\ldots,X_n}(x_1,\ldots,x_n)=f_{X_1}(x_1)\cdots f_{X_n}(x_n)$$

ולהפך. ניתן להוכיח על ידי גזירת פונקצית ההתפלגות של הוקטור.

הוכחה: במקרה הרציף:
ידוע

$$F_{XY}(x,y)=F_x(x)F_y(y)$$

ולכן

$$f_{XY}(x,y)=f_X(x)f_Y(y)$$

כעת נחשב את ההסתברות

$$\begin{aligned}
P\left(X\in A\wedge Y\in B\right) &= \int\int_{A\times B} f_{XY}\left(x,y\right)dxdy\\
&= \int\int_{A\times B} f_X\left(x\right)f_Y\left(y\right)dxdy\\
&= \int_B\left(\int_A f_X\left(x\right)f_Y\left(y\right)dx\right)dy\\
&= \int_B f_Y\left(y\right)\left(\int_A f_X\left(x\right)dx\right)dy\\
&= \int_B f_Y\left(y\right)P\left(X\in A\right)dy\\
&= P\left(X\in A\right)\int_B f_Y\left(y\right)dy\\
&= P\left(X\in A\right)P\left(Y\in B\right)
\end{aligned}$$

■

פונקצית הצפיפות f_{XY} של משתנים בלתי תלויים נראית כמו מלבן, או אוסף מלבנים, מכיוון שבכל תחום שעבורו $f_X>0$ ו-$f_Y>0$ יקבל ערך חיובי.

דוגמאות

1. $X\sim U\left(0,1\right)$, $Y\sim U\left(0,1\right)$ והם בלתי תלויים. איך מתפלג הוקטור $\left(X,Y\right)$?
הצפיפות של X היא

$$\begin{aligned}
f_X\left(x\right) &= 1 \quad 0\le x\le 1\\
f_Y\left(y\right) &= 1 \quad 0\le x\le 1
\end{aligned}$$

$$\begin{aligned}
f_{XY}\left(x,y\right) &= f_X\left(x\right)f_Y\left(y\right)\\
&= 1 \quad 0\le x,y\le 1
\end{aligned}$$

זוהי התפלגות אחידה בריבוע. כלומר ההסתברות ש-$\left(X,Y\right)$ נמצא בתחום D כלשהו היא השטח של $D\cap R$ (כאשר R הוא ריבוע היחידה).

באופן כללי, אם $R\subset\mathbb{R}^2$ תחום חסום כלשהו, אז התפלגות אחידה ב-R פירשה וקטור אקראי $\left(X,Y\right)$ כך שההסתברות שניפול בתחום תחום $D\subset\mathbb{R}^2$ פרופורציונאלית לשטח של $D\cap R$.

2. $R=\left\{x^2+y^2\le 1\right\}$ עיגול היחידה.

$$f_{XY}\left(x,y\right)=\frac{1}{\pi}\qquad\left(x,y\right)\in R$$

$$f_X\left(x\right)=\int_{-\infty}^{\infty}f_{XY}\left(x,y\right)dy=\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}}\frac{1}{\pi}dy=\frac{2\sqrt{1-x^2}}{\pi}\qquad,\quad |x|\le 1$$

$$f_X(x) = 0 \quad , \quad |x| \geq 1$$

המשתנים X ו-Y תלויים: אם נדע מידע כלשהו על X בוודאי שהדבר ישפיע על ההתפלגות של Y, ומעבר לכך אפשר לראות שמכפלת f_X ב-f_y איננה פונקצית הצפיפות f_{XY}.

3. $f_{XY}(x,y) = cx^4y^2 \ , \quad x^2+y^2 \leq 1$ פונקציה זו איננה מתארת משתנים בלתי תלויים, למרות שהיא מכפלה של פונקציה של x בפונקציה של y, מכיוון שהתחום שבו הפונקציה נתמכת איננו מלבני, ובתחום שמחוץ לעיגול היחידה הפונקציה איננה מכפלה של הפונקציות המתאימות.

4. $X_1 \sim \exp(\lambda_1)$, $X_2 \sim \exp(\lambda_2)$ והם ב"ת. מהי ההסתברות ש-X_2 גדול מ-X_1?

$$P(X_2 > X_1) = \int\int_M f_{X_1X_2}(x_1,x_2)\,dx_1dx_2$$

כאשר M הוא המשולש שמעל הקו $x=y$ ברביע הראשון.

$$\begin{aligned}
P(X_2 > X_1) &= \int_0^\infty \int_{x_1}^\infty f_{X_1}(x_1) f_{X_2}(x_2)\,dx_2dx_1 \\
&= \int_0^\infty \int_{x_1}^\infty \lambda_1 e^{-\lambda_1 x_1} \lambda_2 e^{-\lambda_2 x_2}\,dx_2dx_1 \\
&= \lambda_1 \int_0^\infty e^{-\lambda_1 x_1} \int_{x_1}^\infty \lambda_2 e^{-\lambda_2 x_2}\,dx_2dx_1 \\
&= \lambda_1 \int_0^\infty e^{-\lambda_1 x_1} e^{-\lambda_2 x_1}\,dx_1 \\
&= \lambda_1 \int_0^\infty e^{-(\lambda_1+\lambda_2)x_1}\,dx_1 \\
&= \frac{\lambda_1}{\lambda_1+\lambda_2}
\end{aligned}$$

5. נניח ש-X_i ארכי החיים של רכיבים המחוברים בטור. X אורך החיים של המערכת כולה.

נניח ש-$X_i \sim \exp(\lambda)$, ו-X_i בלתי תלויים. מהי ההתפלגות של X?

$$X = \min\{X_i\}$$

$$\begin{aligned}
P(X > x) &= P(\forall i : X_i > x) \\
&= \prod_{i=1}^n P(X_i > x) = \prod_{i=1}^n e^{-\lambda x} = e^{-(\lambda n)x}
\end{aligned}$$

אז פונקציית ההתפלגות של X היא

$$F_X(x) = 1 - P(X > x) = 1 - e^{-(\lambda n)x}$$

כלומר $X \sim \exp(\lambda n)$.

6. $X, Y \sim N(0,1)$, והם בלתי תלויים. נסמן $R = \sqrt{X^2+Y^2}$. מה ההתפלגות של R?

$$\begin{aligned} r>0 \quad , \quad P(R \le r) &= \int\int_{X^2+Y^2 \le r^2} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dxdy \\ &= \frac{1}{2\pi}\int_0^{2\pi}\int_0^r e^{-\frac{\rho^2}{2}} \rho d\rho d\theta \\ &= \int_0^r e^{-\frac{\rho^2}{2}} \rho d\rho \\ &= \left[-e^{-\frac{\rho^2}{2}}\right]_0^r = 1 - e^{-\frac{r^2}{2}} \end{aligned}$$

$$f_R(r) = re^{-\frac{r^2}{2}}, \quad r > 0$$

התפלגות זו נקראת התפלגות ריילי.

7. בתחנת שירות זמן השירות של כל לקוח מתפלג עם צפיפות f_{X_i}. ע"פ התור אני נכנס שלישי. כמה זמן אצטרך לחכות לתורי, מאז פתיחת התחנה?

אם X_i זמן השהות של הלקוח ה-i, וכולנו הגענו לפני זמן הפתיחה, זמן ההמתנה שלי הוא X_1+X_2. נניח גם ש-X_i בלתי תלויים.

$$\begin{aligned} P(X_1+X_2 \le x) &= \int\int_A f_{X_1}(x_1) f_{X_2}(x_2) dx_2 dx_1 \quad , \quad A = \{x_1+x_2 \le x\} \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{x-x_1} f_{X_2}(x_2) f_{X_1}(x_1) dx_2 dx_1 \end{aligned}$$

לחישוב הצפיפות נגזור

$$f_X(x) = \int_{-\infty}^{x} f_{X_1}(x_1) f_{X_2}(x - x_1) dx_1$$

כלומר f_x היא הקונבולוציה של הצפיפויות.

$$f_{X_1+X_2} = f_{X_1} * f_{X_2}$$

כאשר X_1 ו-X_2 בלתי תלויים.

נניח בדוגמה ש-$X_i \sim \exp(\lambda)$

$$\begin{aligned} f_{X_1+X_2}(x) &= \int_0^x \lambda e^{-\lambda x_1} \lambda e^{-\lambda(x-x_1)} dx_1 \\ &= \lambda^2 x e^{-\lambda x} \end{aligned}$$

הגדרה 9.7 התפלגות גמא $X \sim \mathrm{Gamma}\left(r,\lambda\right)$ ניתנת ע"י הצפיפות

$$f_X\left(x\right)=\frac{\lambda^r x^{r-1}e^{-\lambda x}}{\left(r-1\right)!}\quad x\geq 0$$

זוהי הצפיפות של $X=X_1+\cdots+X_r$ כאשר $X_i\sim\exp\left(\lambda\right)$ והם ב"ת.