SGA2 10

Question 1

Assume that continuous random variable X has PDF that is non-zero only on segment $[a, b]$ and strictly positive on open interval (a, b). Median of random variable X is value $m \in (a, b)$ such that $P(X < m) = P(X > m) = 1/2$.

Prove that for symmetrical PDFs median is equal to expected value.

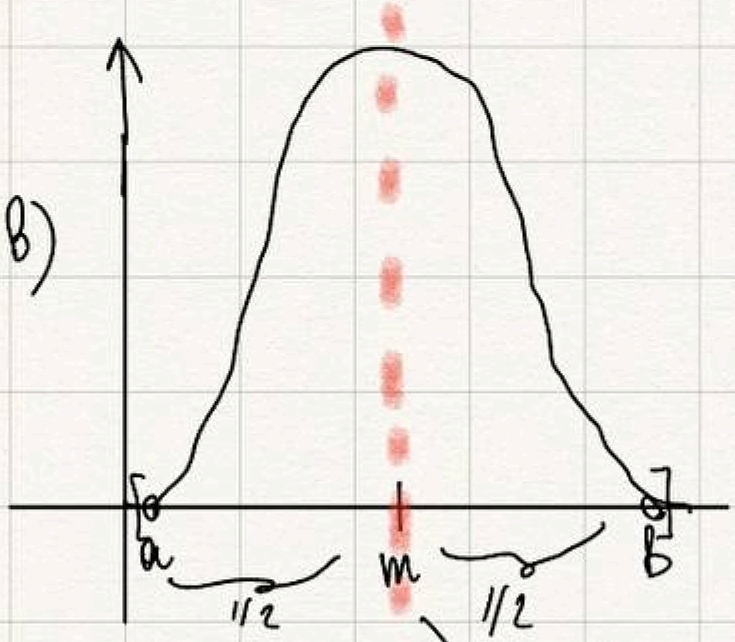
Provide an example of PDF such that median is larger than the expected value.

Let X be random variable, f be strictly increasing function and $Y = f(X)$. What can you say about medians of X and Y?

Solution:

Let X - r.v.

$$f_X(x) = \begin{cases} 0, & x \le a \\ f(x) > 0, & x \in (a,b), \text{ let } m \in (a,b) \\ 0, & x \ge b \end{cases}$$



$P(X<m) = P(X>m) = 1/2$ - median of X

$P(X<m) = CDF_X(m) = \int_{-\infty}^{m} f_X(m)\,dx = \int_a^m f_X(x)\,dx = F(m) - F(a)$ $\to \frac{b+a}{2}$

$P(X>m) = CDF_X(b) - CDF_X(m) = \int_m^b f_X(x)\,dx = F(b) - F(m)$

where $F'(x) = f_X(x)$

Let assume that $f_X(x)$ is symmetrical, i.e. $f\left(\frac{b+a}{2} - t\right) = f\left(\frac{b+a}{2} + t\right)$ for $t \in [0, \frac{b-a}{2}]$, or $f(a+t) = f(b-t)$ for $t \in [0, \frac{b-a}{2}]$

Let's prove that for symmetrical $f_X(x)$, $m = \frac{a+b}{2}$ is median.

$P(X<m) = CDF_X(m) = \int_a^m f(x)\,dx = \int_a^{(a+b)/2} f(x)\,dx = F\left(\frac{a+b}{2}\right) - F(a)$

$P(X>m) = F(b) - F(m) = F(b) - F\left(\frac{a+b}{2}\right) = \int_m^b f(x)\,dx$

Let's replace x with $y = x - m \Rightarrow x = y + m$

$\Rightarrow \int_0^{\frac{b-a}{2}} f(y+m)\,dy$ $(dx = dy)$. Hence $m = \frac{a+b}{2}$ and $f(x)$ is symmetrical to $x = \frac{a+b}{2}$ line then $f(y+m) = f(m-y)$ for $[0; \frac{b-a}{2}]$.

$$P(X>m) = \int_0^{(b-a)/2} f(m-y)\,dy$$

Let's do same replacement in $P(X<m)$ equation.

$$P(X<m) = \int_{-(b-a)/2}^{0} f(m+y)\,dy, \text{ let } z=-y \Rightarrow P(X<m) = \int_{(b-a)/2}^{0} f(m-z)(-dz) =$$

$$= \int_0^{(b-a)/2} f(m-z)\,dz = \int_0^{(b-a)/2} f(m+y)\,dy \text{ by symmetry of } f \Rightarrow$$

$$\Rightarrow P(X<m) = P(X>m) = 1-P(X<m) = P(X<m) = 1/2 \Rightarrow m = \frac{a+b}{2} \text{ is}$$

median.

Let's prove for symmetrical $f_X(x)$ that $E(X) = \frac{a+b}{2} = m$.

$$EX = \int_a^b x \cdot f(x). \text{ Let } y = \frac{a+b}{2} = x-m$$

$$\int_a^b x f(x) = \int_{-(b-a)/2}^{+(b-a)/2} (y+m) f(y+m)\,dy; \; dx = dy$$

$$\int_{-(b-a)/2}^{+(b-a)/2} (y+m) f(y+m)\,dy = \int_{-(b-a)/2}^{(b-a)/2} m \cdot f(y+m)\,dy + \int_{-(b-a)/2}^{(b-a)/2} y f(y+m)\,dy = m\int_a^b f(x)\,dx +$$

$$+ \int_{-(b-a)/2}^{0} y f(y+m)\,dy + \int_0^{(b-a)/2} y f(y+m)\,dy;$$

$\int_a^b f(x)\,dx = 1$ by definition of pdf x. For the 2d interval let's replace y with $z=-y$

$$EX = m \cdot 1 + \int_{+(b-a)/2}^{0} (-z) f(m-z)(-dz); \; (dy = -dz) + \int_0^{(b-a)/2} y f(y+m)\, dy$$

$$EX = m + (-1)\int_0^{\frac{b-a}{2}} z f(m-z)\, dz + \int_0^{(b-a)/2} y f(y+m)\, dy$$ by symmetry of $f(x)$.

$f(m-z) = f(m+z)$ for $z \in [0; \frac{b-a}{2}] \Rightarrow EX = m + \int_0^{(b-a)/2} y f(y+m)\, dy - \int_0^{(b-a)/2} z f(m+z)\, dz$.

It's easy to see that 2 integrals destroy each other $\Rightarrow$

$\Rightarrow EX = m$ - median, which we wanted to prove.

Median $= m = \frac{a+b}{2} = EX$. for symmetrical fx.

Let's show that median can be larger than EX:

$$X \sim pdf_X(x) = \begin{cases} 0, & x \le 0 \\ 2/3, & x \in (0,1) \\ 1/3, & x \in (1,2) \\ 0, & x \ge 2 \end{cases}$$

$$\int_{-\infty}^{\infty} f(x)\, dx = \int_0^1 \frac{2}{3}\, dx + \int_1^2 \frac{1}{3}\, dx = \frac{2}{3} x \Big|_0^1 + \frac{1}{3} x \Big|_1^2 = \frac{2}{3} + \frac{1}{3}(2-1) = 1 \Rightarrow$$

$\Rightarrow f(x)$ is legitimate pdf.

$$EX = \int_{-\infty}^{\infty} x f(x)\, dx = \int_0^1 \frac{2}{3} x\, dx + \int_1^2 \frac{1}{3} x\, dx = \frac{x^2}{3}\Big|_0^1 + \frac{x^2}{6}\Big|_1^2 = \frac{1}{3} + \frac{4}{6} - \frac{1}{6}$$

$= 5/6$

$m = 3/4$ - median of X.

$P(X < m) = \int_0^{3/4} \frac{2}{3} dx = \frac{2}{3} X \Big|_0^{3/4} = \frac{2}{3} \cdot \frac{3}{4} = \frac{1}{2}$

Then let's take $Y = -X \Rightarrow \text{median}_y = -\text{median}_x = -3/4$

$EY = -EX = \frac{-5}{6}$; $\frac{-3}{4} > \frac{-5}{6} \Rightarrow \text{median}_y > EY$

X is r.v., $Y = f(X)$, f - strictly increasing function $\Rightarrow$
$\Rightarrow$ if $P(X < m) = 1/2$ - median_x for some m then
$P(f(X) < f(m)) = 1/2$ because function f preserve the order
i.e. if $a < b \Rightarrow f(a) < f(b)$.

Analogously: $P(f(X) > f(m)) = P(X > m) = 1/2 \Rightarrow f(m)$ is median of Y.

Question 2

Let X be uniform random variable on a segment $[0, 2]$. Consider random variable $Y = X^2$. Find CDF and PDF of Y. Is PDF a bounded function?

Solution:

Let X be Uniform distribution from 0 to 2, i.e. $X \sim \text{Uniform}[0,2]$. Then $PDF_X(x) = \frac{1}{2-0} = \frac{1}{2}$ by the properties of uniform distributions.

Then $CDF_X(x) = \int_{-\infty}^{x} PDF_X(x)dx = \int_0^x \frac{1}{2}dx = \frac{x}{2}\Big|_0^x = \frac{x}{2}, x \in [0,2]$

Let $Y = X^2$ then $CDF_Y(t) = P(Y \le t) = p(X^2 \le t) = p(X \le \sqrt{t}) =$
$CDF_X(\sqrt{t}) = \frac{\sqrt{t}}{2}$ where $\sqrt{t} \in [0,2]$ because $X \in [0,2] \Rightarrow t \in [0^2, 2^2] =$
$= [0,4]$. When $\sqrt{t} \geq 2 \Rightarrow CDF_X(t) = 1$

$PDF_Y(t) = (CDF_Y(t))'$ by definition of PDF (properties of PDF)

$PDF_Y(t) = \left(\frac{\sqrt{t}}{2}\right)' = \frac{1}{4\sqrt{t}}$, $t \in [0,4]$, if $t < 0$ $PDF_Y(t) = 0$

if $t > 4$ $PDF_Y(t) = 0$

$$CDF_Y(t) = \begin{cases} 0, & t < 0 \\ \frac{\sqrt{t}}{2}, & t \in [0,4] \\ 1, & t > 4 \end{cases} \qquad PDF_Y(t) = \begin{cases} 0, & t < 0 \\ \frac{1}{4\sqrt{t}}, & t \in [0,4] \\ 0, & t > 0 \end{cases}$$

Let's look at $PDF_Y(t) = \frac{1}{4\sqrt{t}}$, $t \in [0,4]$

It's strictly decreasing function and continuous on $(0,4] \Rightarrow$
$\Rightarrow PDF_Y(t) \geq PDF_Y(4)$ and $PDF_Y(4) \leftrightarrow \lim_{t \to 0} PDF_Y(0)$;

$PDF_Y(4) = \frac{1}{4\sqrt{4}} = \frac{1}{8}$, $\lim_{t \to 0} PDF_Y(0) = \lim_{t \to 0} \frac{1}{4\sqrt{t}} = +\infty \Rightarrow PDF_Y(t)$ can't be a bounded function. $PDF_Y(t)$ is unbounded.

Question 3

A fair coin is tossed 400 times. Let X be number of heads. Prove that $P(X > 240) \le 1/32$. (Hint: use Chebyshev's inequality and symmetry considerations.)

Solution:

Coin toss is Berulli trial with probability of success $p = 1/2$. Each of $n = 400$ coin tosses is i.i.d.
$Bern(1/2) \Rightarrow X \sim Bern(400, 1/2)$
$EX = np = 400 \cdot 1/2 = 200$ (by properties of EX)
$Var X = np(1-p) = 400 \cdot 1/2 \cdot (1-1/2) = 100$
By Chebyshev's inequality:

$$P(|X - EX| > d) \leq \frac{Var X}{d^2}; \quad \text{Let } d = 40$$

$$P(|X - 200| > 40) \leq \frac{100}{40^2}$$

X can be from 0 to 400 because we can have all n successes at maximum and we can't have less than 0 successes.

$$P(|X - 200| > 40) = P(X < 160 \vee X > 240) \leq \frac{100}{40 \cdot 40} = \frac{1}{16}$$

$$P(X < 160) + P(X > 240) \leq \frac{1}{16}$$

By the symmetry of Bernulli trials with $p = 1/2$ and $160 + 240 = 400$ - number of trials for each elementary outcome when X - number of success, $X > 240$ we can match exactley 1 elementary outcome where $X < 400 - 240 = 160$ by replacing all successes with failures or vice versa. We can do this operation in reverse order so each with $X < 160$ is matched with exactley 1 outcome with $X > 240$. So we bijected all outcomes $X < 160$ to all outcomes $X > 240 \Rightarrow$
$\Rightarrow P(X < 160) = P(X > 240) \Rightarrow P(X < 160) + P(X > 240) = 2P(X > 240) \Rightarrow P(X > 240) \leq \frac{1}{2} \cdot \frac{1}{16} = 1/32$. As we need to prove.

Question 4

Let X and Y be two independent normally distributed random variables with expected value 0 and variance 1. Find their joint PDF. Plot its level curves.

Solution:

SUBJECT NO DATE

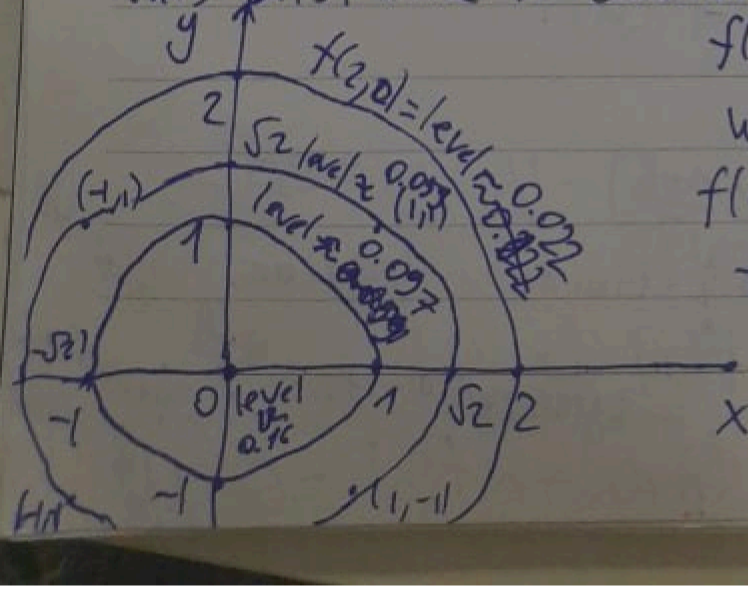
Problem 4

Let $X \sim \mathcal{N}(0,1)$ and $Y \sim \mathcal{N}(0,1)$ which is independent from X.
Then $PDF_X(t) = f_X(t) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{t^2}{2}}$, $PDF_Y(y) = f_y(y) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{y^2}{2}}$.

Because X is independent from Y the Joint $PDF_{X,Y}(x,y) =$
$= f_{x,y}(x,y) = f_X(x) \cdot f_Y(y)$ by the properties of Joint PDF of independent events.

So $f_{x,y}(x,y) = f_X(x) \cdot f_Y(y) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}} \cdot \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{y^2}{2}} =$
$\boxed{= \frac{1}{2\pi} \cdot e^{-\frac{(x^2+y^2)}{2}} \text{ - Joint PDF.}}$

Let's look on Joint PDF level~~s~~ curves.
if $f_{X,Y}(x,y) = c$, c - constant, $c > 0$, then
$c = \frac{1}{2\pi} \cdot e^{-\frac{(x^2+y^2)}{2}} \Rightarrow 2\pi c = e^{-\frac{(x^2+y^2)}{2}} \Rightarrow \ln 2\pi c = -\frac{(x^2+y^2)}{2}$
By π, c - constants, $\ln \pi$, $\ln c$ also are constants as $\ln 2$. $\Rightarrow$
$\Rightarrow \ln 2\pi c = -\frac{(x^2+y^2)}{2} \Rightarrow \ln 2 + \ln \pi + \ln c = (-\frac{1}{2}) \cdot (x^2+y^2) \Rightarrow$
$\Rightarrow x^2+y^2 = (-2) \cdot \ln 2\pi c$, where $(-2) \times \ln 2\pi c$ is some constant d,
where $d \in [\cancel{+\infty}; +\infty)$ because $x^2+y^2 \in [0; +\infty)$ so:
$x^2+y^2 = d = const$ for all x, y which form same level curve, i.e.
$f(x,y) = const.$ ~~for~~
$x^2+y^2 = const$ is equation of circle ~~b~~with center in (0,0) and radius $= r = \sqrt{x^2+y^2}$, because $r = \sqrt{x^2+y^2}$ is distance to point (0,0) and this distance is constant.



$f(1,0) = \frac{1}{2\pi} \cdot e^{-\frac{(1^2+0^2)}{2}} = \frac{1}{2\pi} \cdot e^{-1/2} = \frac{1}{2\pi e^{1/2}} \approx 0.097$

We also have isolated dot (0,0) with it's own level
$f(0,0) = \frac{1}{2\pi} \cdot e^{-\frac{(0^2+0^2)}{2}} = \frac{1}{2\pi} \cdot e^0 = \frac{1}{2\pi} \approx 0.16$
$f(2,0) = \frac{1}{2\pi} \cdot e^{-\frac{2^2+0^2}{2}} = \frac{1}{2\pi} \cdot e^{-2} = \frac{1}{2\pi e^2} \approx 0.022$
$f(1,1) = \frac{1}{2\pi} \cdot e^{-\frac{(1^2+1^2)}{2}} = \frac{1}{2\pi} \cdot e^{-1} = \frac{1}{2\pi e} \approx 0.059$