M378K: September 23^{rd}, 2024.

M378K Introduction to Mathematical Statistics
Problem Set #6
Transformations of Random Variables.

Problem 6.1. *Let X be a continuous random variable with the cumulative distribution function denoted by F_X and the probability density function denoted by f_X.*
Let the random variable $Y = 2X$ have the p.d.f. denoted by f_Y. Then,

(a) $f_Y(x) = 2f_X(2x)$ ✓ ✗

(b) $f_Y(x) = \frac{1}{2}f_X\left(\frac{x}{2}\right)$ ✓

(c) $f_Y(x) = f_X(2x)$ ✗

(d) $f_Y(x) = f_X\left(\frac{x}{2}\right)$ ✗

(e) *None of the above*

$\rightarrow$: $y \in \mathbb{R}$: $F_Y(y) = \mathbb{P}[Y \le y] = \mathbb{P}[2X \le y] =$

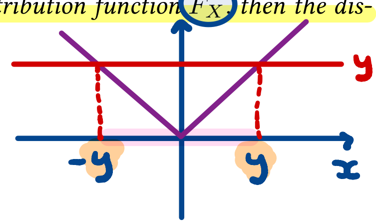
$= \mathbb{P}[X \le \frac{y}{2}] = F_X\left(\frac{y}{2}\right)$

$f_Y(y) = F_Y'(y) = \frac{d}{dy} F_X\left(\frac{y}{2}\right) = \frac{1}{2} f_X\left(\frac{y}{2}\right)$

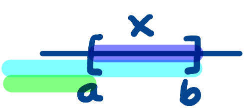
$\square$

Problem 6.2. *If the continuous random variable X has the distribution function F_X, then the distribution function of the random variable $Y = |X|$ equals*

$$F_Y(y) = ?$$



$\rightarrow$: $F_Y(y) = \mathbb{P}[Y \le y]$



$= \mathbb{P}[-y \le X \le y]$

$= \mathbb{P}[X \le y] - \mathbb{P}[X \le -y] = F_X(y) - F_X(-y)$

$y > 0$ $f_Y(y) = (f_X(y) + f_X(-y))$

Remark 6.1. *The goal is to figure out the distribution of the random variable*

$$X = g(Y_1, Y_2, \dots, Y_n)$$

where $Y_i, i = 1, \dots, n$ are a **random sample** *with a common density f_Y.*

1. *Identify the objective: We want* f_X.
2. *Realize:* $f_X = F_X'$
3. *Recall the definition:* $F_X(x) = \mathbb{P}[X \le x] = \mathbb{P}[g(Y_1, \dots, Y_n) \le x]$
4. *Identify the region* A_x *in* $\mathbb{R}^n$ *where*

$$g(\underline{y_1, \dots, y_n}) \le x$$

for every x*, i.e., express*

$$A_x = \{(y_1, \dots, y_n) : g(y_1, \dots, y_n) \le x\}$$

5. *Calculate*

$$F_X(x) = \int \cdots \int_{\mathbb{R}^n} \mathbf{1}_{A_x}(y_1, \dots, y_n) f_Y(y_1) \dots f_Y(y_n) dy_1 \dots dy_n.$$

6. *Differentiate:* $f_X = F_X'$.
7. *Pat yourself on the back!*

Problem 6.3. **One-to-one transformations: Step-by-step** *Let* Y *be a random variable with density* f_Y*. Let* $g : \mathbb{R} \to \mathbb{R}$ *be a strictly increasing differentiable function. Define* $\boxed{\tilde{Y} = g(Y)}$*. What is the density function* $f_{\tilde{Y}}$ *of* $\tilde{Y}$ *expressed in terms of* f_Y *and* g*?*

1. Identify the objective: We want $f_{\tilde{Y}}$.
2. Realize: $f_{\tilde{Y}} = F_{\tilde{Y}}'$
3. Recall the definition:

$$F_{\tilde{Y}}(x) = ?$$

$$F_{\tilde{Y}}(x) = \mathbb{P}[\tilde{Y} \le x] =$$
$$= \mathbb{P}[g(Y) \le x]$$

4. The function g is assumed to be **strictly increasing**. In which way can you modify the inequality in the probability you obtained above to *separate* the random variable Y from the transformation g?

exists $h = g^{-1}$
its inverse function

$$F_{\tilde{Y}}(x) = \mathbb{P}[Y \le h(x)]$$

5. Express your result from above in terms of the c.d.f. F_Y of the r.v. Y.

$$F_{\tilde{Y}}(x) = F_Y(h(x))$$

6. Differentiate: $f_{\tilde{Y}} = F'_{\tilde{Y}}$.

$$f_{\tilde{Y}}(x) = F_{\tilde{Y}}'(x) = \frac{d}{dx} F_Y(h(x)) = f_Y(h(x)) \cdot h'(x)$$

Problem 6.4. *The time T that a manufacturing distribution system is out of operation is modeled by a distribution with the following c.d.f.*

$$F_T(t) = (1 - (2/t)^2)\mathbf{1}_{(2,\infty)}(t) = \begin{cases} 1-4t^{-2} & t>2 \\ 0 & t\le 2 \end{cases}$$

The resulting cost to the company is $Y = T^2$. Find the probability density function f_Y of the r.v. Y.

$\rightarrow$: $g(x)=x^2,\ x>2 \Rightarrow h(x)=\sqrt{x},\ x>4 \Rightarrow h'(x)=\frac{1}{2\sqrt{x}}$

$$f_T(t) = \frac{8}{t^3} 1_{(2,\infty)}(t)$$

$$f_Y(y) = \frac{1}{2\sqrt{y}} \cdot \frac{8}{(\sqrt{y})^3} 1_{(y>4)} = \frac{4}{y^2} 1_{(4,\infty)}(y)$$

Problem 6.5. *What if h is strictly decreasing?* $h = g^{-1}$

$$F_{\tilde{Y}}(y) = \mathbb{P}[\tilde{Y} \le y] = \mathbb{P}[g(Y) \le y] = \mathbb{P}[Y \ge h(y)] = 1 - \mathbb{P}[Y < h(y)]$$

$$f_{\tilde{Y}}(y) = -f_Y(h(y))h'(y)$$

$x_1 < x_2 \Rightarrow h(x_1) > h(x_2)$

$\uparrow$

decreasing for all x_1, x_2

Problem 6.6. *The unifying formula?*

$$f_{\tilde{Y}}(y) = f_Y(h(y)) \cdot |h'(y)|$$

Do not forget: it always makes sense to simply attack a problem without giving it a “label” …. Just look at the following problem:

Problem 6.7. *Let T_1 and T_2 be independent geometric random variables with parameters $p_1 = 1/2$ and $p_2 = 1/3$. Compute $\mathbb{E}[\min(T_1, T_2)]$.*