Universität Bielefeld
Tanja Pasurek
Ehsan Abedi WiSe 2023/24

Foundations of Statistics

Homework 6

Exercise 1 (Transformed density functions).

Let X be a continuous random variable. Define a new random variable $Y := g(X)$, where g is some map. In this exercise, we answer the following **important question:** What is the distribution of Y?

(a) Transformation formula in univariate case:

Let $X : \Omega \to \mathcal{I} \subseteq \mathbb{R}$ be a continuous random variable taking values in some interval $\mathcal{I}$. Suppose that the map $g : \mathcal{I} \to \mathbb{R}$ is strictly monotone increasing or strictly monotone decreasing, so that it has an inverse $h = g^{-1} : \mathcal{J} \to \mathcal{I}$ defined on the image set $\mathcal{J} := g(\mathcal{I}) = \{y \in \mathbb{R} : y = g(x) \text{ for some } x \in \mathcal{I}\}$. Moreover, assume that there exist a continuous derivative $g'(x) \neq 0$ for all $x \in \mathcal{I}$, which in turn guarantees the existence of

$$h'(y) = \frac{\partial}{\partial y} g^{-1}(y) = \frac{1}{g'(h(y))} \quad \text{for all } y \in \mathcal{J}.$$

Then show that the continuous random variable $Y := g(X)$ has PDF

$$f_Y(y) = \left|h'(y)\right| \cdot f_X(h(y)), \quad y \in \mathcal{J}. \tag{1}$$

Hint: use the so-called **CDF method**.

Apply formula (1) for the case of linear transformations and compare it with we found in Exercise 4(b), HW 4.

(b) Formula (1) can be generalized to multivariate case:

Let $\mathbf{X} : \Omega \to \boldsymbol{\mathcal{X}} \subseteq \mathbb{R}^n$ be a continuous random variable with joint density function $f_{\mathbf{X}}$. Let $g : \mathcal{X} \to \mathbb{R}^n$ be differentiable bijection with non-zero derivative. Prove that $\mathbf{Y} := g(\mathbf{X})$ has joint density function

$$f_{\mathbf{Y}}(\boldsymbol{y}) = \left|\det(J_{g^{-1}}(\boldsymbol{y}))\right| \cdot f_{\mathbf{X}}(g^{-1}(\boldsymbol{y})), \qquad \boldsymbol{y} \in \boldsymbol{\mathcal{Y}} := g(\boldsymbol{\mathcal{X}}), \tag{2}$$

whenever $J_{g^{-1}}(\boldsymbol{y})$ is well-defined. Here $J_{g^{-1}}(\boldsymbol{y})$ is the Jacobian matrix (i.e. the matrix of partial derivatives) of the map g^{-1}.

Hint: For each (Borel) subset $A \subset \mathbb{R}^n$, find $\mathbb{P}(\boldsymbol{Y} \in A)$ and apply the change of variables formula as in *Calculus I*.

Exercise 2 (Transformed density functions, examples in 1D).

(a) Let a random variable X have PDF $f_X(x) = e^{-x}$ for $x > 0$. Define $Y := g(X) = \log X$. Using the above scheme, check that

$$f_Y(y) = e^y e^{-e^y} \quad \text{for } y \in \mathbb{R}.$$

(*Warning:* Although $Y := g(X)$, in general $f_Y \neq g(f_X)$!)

(b) Let $X \sim \mathrm{Unif}(0,1)$. Find the distribution of the random variable $Y = X^2$. Check your answer using simulation in `R`. To this end, simulate a large number (for instance, $n = 10^5$) of samples from the uniform distribution, square the values, make a histogram (with `freq=FALSE`) and superimpose the calculated density on top of the histogram. Compute $\mathbb{E}[Y]$ using both the LOTUS rule and the density f_Y that you have found.

Exercise 3 (Transformed density functions, examples in 2D).

(a) Let X and Y be independent, continuous random variables with densities f_X and f_Y. Use formula (2) to find the density of $X + Y$ and compare the result with what we found in Exercise 2, HW 5.

(b) Let $X, Y \sim N(0,1)$ be independent. Show that $X+Y$ is independent of $X - Y$. (*Hint:* define $U := X + Y$ and $V := X - Y$ and compute their joint density $f_{U,V}(u,v)$.)

(c) Let $X_1, X_2 \sim N(0,1)$ be independent. Write the sample mean $\overline{X}$ and the sample variance S^2 in terms of X_1, X_2. Are $\overline{X}$ and S^2 independent or not?

Exercise 4 (A universal random number generator, cf. Ch. 1.9).

Let $X : \Omega \to \mathbb{R}$ be a continuous random variable. Suppose that its CDF F_X is continuous and strictly increasing from 0 to 1 on some interval $\mathcal{I} \subseteq \mathbb{R}$. In this case, F_X has an inverse function $F_X^{-1} : [0,1] \to \mathcal{I}$.

(a) Define $Y := F_X(X)$, i.e., you *plug a continuous random variable into its own CDF*. Show that $Y \sim \mathrm{Unif}(0,1)$. This is called the **probability integral transform**.

(b) Let now $U \sim \mathrm{Unif}(0,1)$ and define $Z := F_X^{-1}(U)$, i.e., you *plug a uniform random variable into an inverse CDF*. Show that Z and X have the same distribution, i.e., $F_Z = F_X$.

▶ **Conclusion:** *Any continuous real-valued random variable can be transformed into a uniform random variable and back by using its CDF.*

(c) Write an `R` code to simulate continuous random variables from the density

$$f(x) = \frac{2}{(x+1)^3}, \quad x > 0.$$

Make a histogram of $n = 10^5$ simulated values and superimpose the density function to check the work.

Hint: The distribution is heavy-tailed, so in order to make a nice histogram, plot only the values less than 10 (which is about 99% of the values).

Exercise 5.

(a) Assume that $X : \Omega \to \mathbb{R}$ and $Y : \Omega \to \mathbb{R}$ are independent random variables, with X having a continuous distribution. Show that

$$\mathbb{P}\Big(\{\omega \in \Omega : X(\omega) \neq Y(\omega)\}\Big) = 1.$$

(b) Let X_n, $n \geq 1$, be a sequence of independent continuous random variables. Show that

$$\mathbb{P}\Big(\{\omega \in \Omega : X_i(\omega) = X_j(\omega) \text{ for some distinct indexes } i, j \geq 1\}\Big) = 0.$$

(c) (From discrete to continuous uniform). Let X_n, $n \geq 1$, be a discrete random variable taking values in $\left\{\frac{1}{n+1}, ..., \frac{n}{n+1}\right\}$ uniformly. Show that X_n converges in distribution to the uniform distribution on $[0, 1]$.

Exercise 6 (Simulation of Law of Large Numbers (LLN) in `R`).

(a) To begin with, plot the densities of normal distribution with mean 2 and variance 1 (in blue) and Cauchy distribution with location parameter 2 and scale parameter 1 (in red) on the same plot. Which one has a heavier tail?

(b) Take a sample of $n = 5000$ realizations from $N(2, 1)$. Calculate the cumulative arithmetic mean of your sample, that is the arithmetic mean of the first number, of the first two numbers, and so on (see ?`cumsum`). Plot the mean values obtained and overlaid them with a horisontal line corresponding to the actual mean value.

(c) Repeat the procedure in (b) with the Cauchy distribution with with location parameter 2 and scale parameter 1 (see ?`rcauchy`). Can we observe a similar convergence in this case? Justify your answer.

Hint: if you want to get a reproducible sequence of random numbers, use the command `set.seed` to start a random generator with any number of your choice (see e.g. [*Heumann et al.*], Appendix: Introduction to `R`, p. 418).

Exercise 7 (Simulation of Central Limit Theorem (CLT) in `R`).

Let $X_i, i \geq 1$ be a sequence of i.i.d. random variables with $\mathbb{E}[X_i] = \mu$ and $\mathrm{Var}(X_i) = \sigma^2$ such that $\sigma^2 \in (0, \infty)$. CLT tells us that the distribution of standardized sum

$$Z_n := \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\sigma},$$

converges to the standard normal distribution $N(0, 1)$.

To see this, you need to consider two natural numbers k and n. Fix $k = 1000$. At first, take an arbitrary n and generate $k = 1000$ random samples Z_n when we have i.i.d. Pois(0.5)–distributed random variables X_i, $1 \leq i \leq n$. Plot the corresponding histogram and overlaid it with the density of the normal distribution. Then increase n, while keep k fixed. Repeat the simulation when we have i.i.d. Exp(2)–distributed random variables X_i, $1 \leq i \leq n$. Does the result depend on distribution of X_i?