Course:	Mathematical Statistics
Term:	Fall 2017
Instructor:	Gordan Žitković

Lecture 4
Functions of random variables

Let Y be a random variable, discrete and continuous, and let g be a function from $\mathbb{R}$ to $\mathbb{R}$, which we think of as a transformation. For example, Y could be a height of a randomly chosen person in a given population in inches, and g could be a function which transforms inches to centimeters, i.e. $g(y) = 2.54 \times y$. Then $W = g(Y)$ is also a random variable, but its distribution (pdf), mean, variance, etc. will differ from that of Y. Transformations of random variables play a central role in statistics, and we will learn how to work with them in this section.

4.1 Computing expectations

Expectations of functions of random variables are easy to compute, thanks to the following result, sometimes known as the *fundamental formula*.

Theorem 4.1.1. *Suppose that Y is a random variable, g is a* ***transformation****, i.e., a real function, and $W = g(Y)$. Then*

1. *if Y is discrete, with pmf p_Y, we have*

$$\mathbb{E}[W] = \sum_{y \in \mathcal{S}_Y} g(y)\, p_Y(y),$$

2. *if Y is continuous, with pdf f_Y, we have*

$$\mathbb{E}[W] = \int_{-\infty}^{\infty} g(y) f_Y(y)\, dy.$$

We have already used this formula, without knowing, when we wrote down a formula for the variance

$$\operatorname{Var}[Y] = \mathbb{E}[(Y - \mu_Y)^2] = \int_{-\infty}^{\infty} (y - \mu_Y)^2 f_Y(y)\, dy.$$

Indeed, we applied the transformation $g(y) = (y - \mu_Y)^2$ to Y and then computed the expectation of the new random variable $W = g(Y) = (Y - \mu_Y)^2$.

Example 4.1.2. The stopping distance (the distance traveled by the car from the moment the brake is applied to the moment it stops) in feet is a quadratic function of the car's speed (in mph), i.e.,

$$g(y) = cy^2$$

where c is a constant which depends on the physical characteristics of the car, its brakes, the road surface, etc. For the purposes of this example, let's take a realistic value $c = 0.07$ (in appropriate units).

In a certain traffic study, the distribution of cars' speeds at the onset of breaking is empirically determined to be uniformly distributed on the interval $[60, 90]$, measured in miles per hour. What is the expected value of the stopping distance?

The stopping distance W is given by $W = g(Y)$, where $g(y) = 0.07\,y^2$, and so, according to our formula, we have

$$\begin{aligned} \mathbb{E}[W] &= \int_{-\infty}^{\infty} g(y) f_Y(y)\,dy = \int_{-\infty}^{\infty} 0.07\,y^2 \tfrac{1}{90-60} \mathbf{1}_{[60,90]}(y)\,dy \\ &= 0.07/30 \int_{60}^{90} y^2\,dy = 399. \end{aligned}$$

If we compute the expected speed, we get

$$\mathbb{E}[Y] = \tfrac{1}{30} \int_{60}^{90} y\,dy = 75,$$

and if we compute the stopping distance of the car traveling at 75 mph, we get

$$g(75) = 393.75.$$

It follows that the average (expected) stopping distance is not the same as the stopping distance corresponding to the average speed. Why is that?

Caveat: What we observed at the end of Example 4.1.2 is so important that it should be repeated:

$$\text{In general, } \mathbb{E}[g(Y)] \neq g(\mathbb{E}[Y])!.$$

In fact, the only time we can guarantee equality for any Y is when g is an affine function, i.e., when $g(y) = \alpha y + \beta$ for some constants α and β.

4.2 The cdf-method

The fundamental formula of Theorem 4.1.1 is useful for computing expectations, but it has nothing to say about the *distribution* of $W = g(Y)$. For example, we may wonder whether the distribution of stopping distances is uniform on some interval, just like the distribution of velocities at the onset of breaking in Example 4.1.2. There are several methods for answering this question, and we start with the one which almost always works - the **cdf-method**.

Suppose that we know the cdf F_Y of Y and that we we are interested in the distribution of $W = g(Y)$. Using the definition of the cdf F_W of W, we can write

$$F_W(w) = \mathbb{P}[W \leq w] = \mathbb{P}[g(Y) \leq w].$$

The probability on the right is not quite the cdf of Y, but if it can be rewritten *in terms* of probabilities involving Y, or, better, the cdf of Y, we are in business:

1. If g is strictly increasing, then it admits an inverse function g^{-1} and we can write

   $$F_W(w) = \mathbb{P}[g(Y) \leq w] = \mathbb{P}[Y \leq g^{-1}(w)] = F_Y(g^{-1}(w)),$$

   and we have an expression of F_W in terms of F_Y. Once F_W is known, it can be used further to compute the pdf (in the continuous case) or the pmf (in the discrete case), or ...

2. A very similar computation can be made if g is strictly decreasing. The only difference is that now $\mathbb{P}[g(Y) \leq w] = \mathbb{P}[Y \geq g^{-1}(w)]$. In the continuous case we have $\mathbb{P}[Y \geq y] = 1 - F_Y(y)$ (why only in continuous?), so

   $$F_W(w) = \mathbb{P}[g(Y) \leq w] = \mathbb{P}[Y \geq g^{-1}(w)] = 1 - F_Y(g^{-1}(w)).$$

3. The function g is neither increasing nor decreasing, but the inequality $g(y) \leq w$ can be "solved" in simple terms. To understand what is meant by this, have a look at examples below.

Example 4.2.1.

1. **Linear transformations.** Let Y be any random variable, and let $W = g(Y)$ where $g(y) = a + by$ is a linear transformation with $b > 0$. Since $b > 0$, the function g is strictly increasing. Therefore,

   $$\begin{aligned} F_W(w) &= \mathbb{P}[g(Y) \leq w] = \mathbb{P}[a + bY \leq w] = \mathbb{P}[Y \leq \tfrac{w-a}{b}] \\ &= F_Y(\tfrac{w-a}{b}). \end{aligned}$$

   This expression is especially nice if Y is a continuous random variable because then so is W, and we have

$$f_W(w) = \tfrac{d}{dw} F_W(w) = \tfrac{d}{dw} F_Y(\tfrac{w-a}{b}) = f_Y(\tfrac{w-a}{b})\tfrac{1}{b},$$

where the last equality follows from the chain rule. Here are some important special cases:

a) **Linear transformations of a normal.** If $Y \sim N(0,1)$ is a unit normal, then $f_Y(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2}$, and so,

$$f_W(w) = \frac{1}{\sqrt{2\pi b^2}} e^{-\frac{(w-a)^2}{2b^2}}.$$

We recognize this as the pdf of the normal distribution, but this time with paramters a and b. This inverse of this computation lies behind the familiar z**-score** transformation: if $Y \sim N(\mu, \sigma)$, then $Z = \frac{Y-\mu}{\sigma} \sim N(0,1)$.

b) **Linear transformations of a uniform.** If $Y \sim U(0,1)$ is a random number, $g(y) = a + by$ and $W = g(Y)$, then $F_W(w) = F_Y((w-a)/b)$ and, so,

$$f_W(w) = \tfrac{1}{b} f_Y((w-a)/b)) = \tfrac{1}{b}\mathbf{1}_{[0,1]}((w-a)/b).$$

When we talked about indicators, we mentioned that a different notation for the same function can simplify computations in some cases. Here is the case in point. If we replace $\mathbf{1}_{[0,1]}((w-a)/b)$ by $\mathbf{1}_{\{0 \leq (w-a)/b \leq 1\}}$, we can rearrange the expression inside $\{\}$ and get

$$f_W(w) = \tfrac{1}{b}\mathbf{1}_{\{a \leq w \leq a+b\}} = \tfrac{1}{b}\mathbf{1}_{[a,a+b]}(w),$$

and we readily recognize f_W as the pdf of another uniform distribution, but this time with parameters a and b. If we wanted to transform $U(0,1)$ into $U(l,r)$, we would simply need to pick $a = l$ and $b = r - l$.

It is not a coincidence that *linear* transformations of normal and uniform random variables result in random variables in the same parametric families (albeit with different parameters). Parametric families are often (but not always) chosen to have this exact property.

2. **Inverse Exponential distribution.** Let $Y \sim E(\tau)$ be an exponentially distributed random variable, and let $g(y) = 1/y$. The function g is strictly decreasing on $(0, \infty)$ and so, for $w > 0$, we have

$$F_W(w) = \mathbb{P}[1/Y \leq w] = \mathbb{P}[Y \geq 1/w] = 1 - F_Y(1/w) = e^{-\frac{1}{\tau w}}.$$

This computation will not work for $w \leq 0$, but we know that W always takes positive values, as it is the reciprocal of Y, which is always positive. Therefore, $F_W(w) = 0$ for $w \leq 0$. We can differentiate the expression for F_W to obtain the pdf f_W:

$$f_W(w) = e^{-\frac{1}{\tau w}} \frac{1}{\tau w^2} \mathbf{1}_{(0,\infty)}(w).$$

This pdf cannot be recognized as the pdf of any of our named distributions, but it is sometimes called the **inverse exponential distribution**, and it is used in wireless communications.

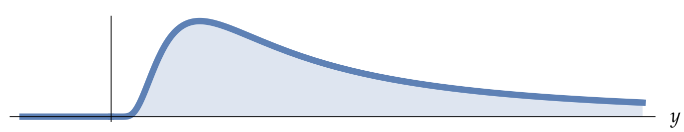

Figure 1. The pdf of the inverse exponential distribution.

3. χ^2**-distribution.** Let $Y \sim N(0,1)$ be the unit normal random variable, and let $g(y) = y^2$. This is an example of a transformation which is neither increasing nor decreasing. We can still try to make sense of the expression $g(y) \leq w$, i.e., $y^2 \leq w$ and use the cdf-method:

$$F_W(w) = \mathbb{P}[Y^2 \leq w].$$

For $w < 0$ it is impossible that $Y^2 < w$, so we immediately conclude that $F_W(w) = 0$ for $w < 0$. When $w \geq 0$, we have

$$\mathbb{P}[Y^2 \leq w] = \mathbb{P}[Y \in [-\sqrt{w}, \sqrt{w}]] = \mathbb{P}[Y \leq \sqrt{w}] - \mathbb{P}[Y < -\sqrt{w}].$$

Since $\mathbb{P}[Y = -\sqrt{w}] = 0$ (as Y is continuous), we get

$$F_W(w) = \begin{cases} 0, & w < 0 \\ F_Y(\sqrt{w}) - F_Y(-\sqrt{w}), & w \geq 0. \end{cases}$$

We differentiate both sides in w and use the fact that F_Y is a normal cdf so that $\frac{d}{dy} F_Y(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2}$ to obtain

$$f_W(w) = \frac{1}{\sqrt{2\pi w}} e^{-\frac{1}{2}w} \mathbf{1}_{(0,\infty)}(w). \tag{4.2.1}$$

A random variable with the pdf $f_W(w)$ of (4.2.1) above is said to have a χ^2**-distribution** (pronounced [kai-skwer]). It is very important in statistics, and we will spend a lot more space on it later.

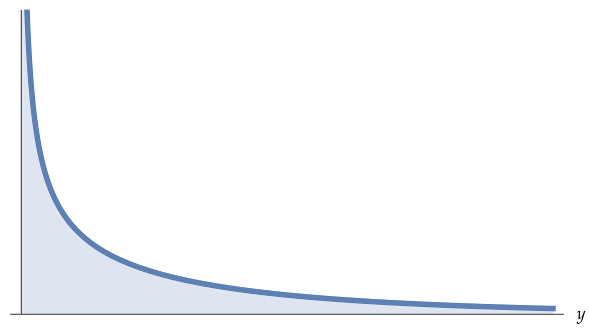

Figure 2. The pdf of the χ^2 distribution.

4.3 The h-method

The application of the cdf-method can sometimes be streamlined, leading to the so-called h**-method** or the **method of transformations**. It works when Y is a continuous random variable and when the transformation function g admits an inverse function h. Supposing that is the case, remember that, when g is increasing, we have

$$F_W(w) = F_Y(g^{-1}(w)) = F_Y(h(w)).$$

If we assume that everything is differentiable and that Y and W admit pdfs f_Y and f_W, we can take a derivative in w to obtain

$$f_W(w) = \tfrac{d}{dw}F_W(w) = \tfrac{d}{dw}\Big(F_Y(h(w))\Big) = \tfrac{d}{dw}F_Y(h(w))\tfrac{d}{dw}h(w) = f_Y(h(w))h'(w).$$

Another way of deriving the same formula is to interpret the pdf $f_Y(y)$ as the quantity such that

$$\mathbb{P}[Y \in [y, y + \Delta y]] \approx f_Y(y)\,\Delta y,$$

when $\Delta y > 0$ is "small". Applying the same to $W = g(Y)$ in two ways yields

$$\mathbb{P}[W \in [w, w + \Delta w]] \approx f_W(w)\,\Delta w,$$

by also, assuming that g is increasing,

$$\begin{aligned}\mathbb{P}[W \in [w, w + \Delta w] &= \mathbb{P}[g(Y) \in [w, w + \Delta w]] = \mathbb{P}[Y \in [h(w), h(w + \Delta w)]] \\ &\approx \mathbb{P}[Y \in [h(w), h(w) + \Delta w\, h'(w)]] = f_Y(h(w))h'(w)\Delta w.\end{aligned}$$

The approximate equality $\approx$ between the third and the fourth probability above uses the fact that $h(w + \Delta w) \approx h(w) + \Delta w \, h'(w)$ which is nothing but a consequence of the definition of the derivative

$$h'(w) = \lim_{\Delta w \to 0} \tfrac{h(w+\Delta w) - h(w)}{\Delta w}.$$

It could also be seen as the first-order Taylor formula for h around w.

The derivation above can be made fully rigorous, leading to the following theorem (why does the absolute value $|h'(w)|$ appear there?). The word *interval* in it means (a, b), where either a or b could be infinite (so that, for example, $\mathbb{R}$ itself is also an interval).

Theorem 4.3.1 (The h-method)**.** *Suppose that the function g is*

1. *defined on an interval $I \subseteq \mathbb{R}$.*
2. *its image is an interval $J \subseteq \mathbb{R}$*
3. *g has a continuously-differentiable inverse function $h : J \to I$*

Suppose that Y is a continuous random variable with pdf f_Y such that $f_Y(y) = 0$ for $y \notin I$. Then $W = g(Y)$ is also a continuous random variable and its pdf is given by the following formula:

$$f_W(w) = f_Y(h(w)) \left| h'(w) \right| \mathbf{1}_{\{w \in J\}}.$$

Note: in almost all applications $I = \{y \in \mathbb{R} : f_Y(y) > 0\}$, for a properly defined version of f_Y and $J = g(I)$.

Example 4.3.2.

1. Let Y be a continuous random variable with pdf

   $$f_Y(y) = \tfrac{1}{\pi(1+y^2)}.$$

   The distribution of Y is called the **Cauchy distribution**. We define $W = g(Y)$ where $g(y) = \arctan(y)$. The function g is defined on the interval $I = \mathbb{R}$ and its image is the interval $J = (-\pi/2, \pi/2)$. Moreover, its inverse is the function $h : J \to I$ given by

   $$h(w) = \tan(w).$$

   This function admits a derivative $h'(w) = 1/\cos^2(w)$.

   Theorem 4.3.1 can be applied and it states that

   $$f_W(w) = \tfrac{1}{\pi(1+\tan^2(w))} \tfrac{1}{\cos^2(w)} \mathbf{1}_{\{w \in (-\pi/2, \pi/2)\}} = \tfrac{1}{\pi} \mathbf{1}_{\{w \in (-\pi/2, \pi/2)\}}.$$

This allows us to identify the distribution of W as *uniform* on the interval $(-\pi/2, \pi/2)$, i.e., $Y \sim U(-\pi/2, \pi/2)$.

2. Let $Y \sim E(\tau)$, and let $g(y) = \sqrt{y}$. The function g is defined on $I = (0, \infty)$ and maps it into $J = (0, \infty)$, and its inverse is $h(w) = w^2 : J \to I$. The pdf f_Y of y is given by

$$f_Y(y) = \tfrac{1}{\tau}\exp(-y/\tau)\mathbf{1}_{\{y>0\}}$$

and, so, by Theorem 4.3.1 $W = \sqrt{Y} = g(Y)$ is a continuous random variable with density

$$f_W(w) = \tfrac{2}{\tau}\, w \exp(-w^2/\tau)\mathbf{1}_{\{w>0\}},$$

where we removed the absolute value around $h'(w) = 2w$ because of the indicator $\mathbf{1}_{w>0}$. This is known as the **Weibull distribution**.

3. Let Y and W be as in Example 4.1.2, i.e., $Y \sim U(60, 90)$ and $g(y) = cy^2$, where $c = 0.07$. Then $g : \mathbb{R} \to \mathbb{R}$ is neither increasing nor decreasing, and does not admit an inverse. However, it is increasing on the set $(60, 90)$ where the random variable Y takes place, i.e., where $f_Y(y) > 0$ (we can exclude the end-points 60 and 90 because they happen with probability 0). If we restrict g to the interval $I = (60, 90)$, it admits an inverse $h : J \to I$, where $J = (g(60), g(90)) = (252, 567)$, and

$$h(w) = \sqrt{\tfrac{w}{c}}, w \in J.$$

The pdf $f_W(w)$ of W is then given by

$$f_W(y) = f_Y(h(w))h'(w) = \tfrac{1}{90-60}\tfrac{1}{2\sqrt{c}}w^{-1/2}\mathbf{1}_{\{w\in J\}}.$$

Here are the graphs of the pdfs of Y and $W = g(Y)$:

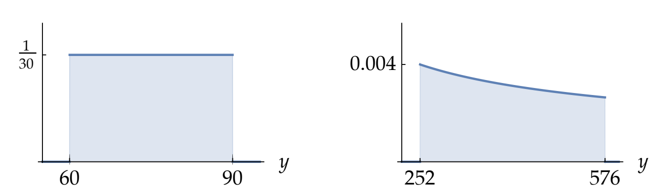

Figure 3. The pdfs of Y and $W = g(Y)$ (with both axes scaled differently)

4.4 Problems

Problem 4.4.1. Let Y be an exponential random variable with parameter $\tau > 0$. Compute the cdf F_W and the pdf f_W of the random variable $W = Y^3$.

Problem 4.4.2. Let Y be a uniformly distributed random variable on the interval $[0, 1]$, and let $W = \exp(Y)$. Compute $\mathbb{E}[W]$, the CDF F_W of W and the pdf f_W of W.

Problem 4.4.3. A scientist measures the side of cubical box and the result is a random variable (due to the measurement error) Y, which we assume is normally distributed with mean $\mu = 1$ and $\sigma = 0.1$ (both in feet). In other words, the true measurement of the side of the box is 1 foot, but the scientist does not know that; she only knows the value of Y.

1. What is the distribution of the volume W of the box?
2. What is the probability that the scientist's measurement overestimates the volume of the box by more than 10%? (*Hint:* Review z-scores and computations of probabilities related to the normal distribution. You will not need to integrate anything here, but may need to use software (if you know how) or look into a normal table. We will talk more about how to do that later. For now, use any method you want.)

Problem 4.4.4. Let Y be a continuous random variable with the density function f_Y given by

$$f_Y(y) = \begin{cases} \frac{2+3y}{16}, & y \in (1, 3) \\ 0, & \text{otherwise.} \end{cases}$$

The pdf $f_W(w)$ of $W = Y^2$ is

(a) $f_W(w) = \frac{2+3\sqrt{w}}{16\sqrt{w}} \mathbf{1}_{\{w \in (1,3)\}}$

(b) $f_W(w) = \frac{2+3\sqrt{w}}{32\sqrt{w}} \mathbf{1}_{\{w \in (1,9)\}}$

(c) $f_W(w) = \frac{2+3w}{16\sqrt{w}} \mathbf{1}_{\{w \in (1,9)\}}$

(d) $f_W(w) = \frac{2+3w}{32\sqrt{w}} \mathbf{1}_{\{w \in (1,3)\}}$

(e) none of the above

Problem 4.4.5. The **Maxwell-Boltzmann** distribution describes the distribution of speeds of particles in an (idealized) gas, and its pdf is given by

$$f_Y(y) = \frac{4\beta^3}{\sqrt{\pi}} y^2 e^{-\beta y^2} \mathbf{1}_{(0,\infty)}(y),$$

where $\beta > 0$ is a constant that depends on the properties of the particular gas studied.

The *kinetic energy* of a gas particle with speed y is $\frac{1}{2}my^2$. What is the distribution (pdf) of the kinetic energy if the particle speed follows the Maxwell-Boltzmann distribution?

Problem 4.4.6. Let Y be a uniform distribution on $(0,1)$. The distribution of the random variable $-\frac{1}{2}\log(Y)$ is

(a) exponential $E(\tau)$ with $\tau = 2$

(b) exponential $E(\tau)$ with $\tau = 1/2$

(c) uniform $U(0,1/2)$ on $(0,1/2)$

(d) uniform $U(0,2)$ on $(0,2)$

(e) none of the above

Problem 4.4.7. Let Y be an exponential random variable with parameter $\tau > -1$. Then $\mathbb{E}[e^{-Y}] =$

(a) τ (b) $\frac{1}{1+\tau}$ (c) $\frac{1}{1-\tau}$ (d) $\frac{\tau}{1-\tau}$ (e) none of the above

Problem 4.4.8. Let Y be a random variable with the pdf $f_Y(y) = \frac{1}{\pi(1+y^2)}$. The pdf of $W = 1/Y^2$ is

(a) $\frac{2}{\pi\sqrt{w}(1+w)}\mathbf{1}_{\{w>0\}}$

(b) $\frac{1}{\pi\sqrt{w}(1+w)}\mathbf{1}_{\{w>0\}}$

(c) $\frac{w^2}{\pi(1+w^4)}$

(d) $\frac{2w^2}{\pi(1+w^4)}$

(e) none of the above

(*Hint:* You do not need to actually compute the cdf of Y.)

Problem 4.4.9. The pdf of $W = 1/Y^2$, where $Y \sim E(\tau)$ is

(a) $\frac{2}{\tau}y^{-3/2}e^{-y/\tau}\mathbf{1}_{\{y>0\}}$

(b) $\frac{1}{2\tau}(-y^{-3/2})e^{-\sqrt{y}/\tau}\mathbf{1}_{\{y>0\}}$

(c) $\frac{1}{\tau}e^{-1/(y^2\tau)}\mathbf{1}_{\{y>0\}}$

(d) $\frac{1}{2\tau y^{3/2}}e^{-1/(\tau\sqrt{y})}\mathbf{1}_{\{y>0\}}$

(e) none of the above

Problem 4.4.10. Let Y be a uniform random variable on $[-1,1]$, and let $W = Y^2$. The pdf of W is

(a) $\frac{1}{4\sqrt{|w|}}\mathbf{1}_{\{-1<w<1\}}$

(b) $\frac{1}{\sqrt{w}}\mathbf{1}_{\{0<w<1\}}$

(c) $\frac{1}{2\sqrt{w}}\mathbf{1}_{\{0<w<1\}}$

(d) $2w\mathbf{1}_{\{0<w<1\}}$

(e) none of the above

Problem 4.4.11. Let Y be a uniform random variable on $[0,1]$, and let $W = Y^2$. The pdf of W is

(a) $\frac{1}{2\sqrt{|w|}}\mathbf{1}_{\{-1<w<1\}}$

(b) $\frac{1}{\sqrt{w}}\mathbf{1}_{\{0<w<1\}}$

(c) $\frac{1}{2\sqrt{w}}\mathbf{1}_{\{0<w<1\}}$

(d) $2w\mathbf{1}_{\{0<w<1\}}$

(e) none of the above

Problem 4.4.12. Fuel efficiency of a sample of cars has been measured by a group of American engineers, and it turns out that the distribution of gas-mileage is uniformly distributed on the interval [10 mpg, 30 mpg], where mpg stands for *miles per gallon*. In particular, the average gas-mileage is 20 mpg.

A group of European engineers decided to redo the statistics, but, being European, they used European units. Instead of *miles* per *gallon*, they used *liters* per 100 *km*. (Note that the ratio is reversed in Europe - higher numbers in liters per km correspond to *worse* gas mileage). If one mile is 1.609 km and one gallon is 3.785 liters, the average gas-mileage of 20 mpg, obtained by the Americans, translates into 11.762 liters per 100 km. However, the average gas mileage obtained by the Europeans was different from 11.762, even though they used the same sample, and made no computational errors. How can that be? What did they get? Is the distribution of gas-mileage still uniform, when expressed in European units? If not, what is its pdf?

Problem 4.4.13. Let Y be a uniformly distributed random variable on the interval $(0,1)$. Find a function g such that the random variable $W = g(Y)$ has the exponential distribution with parameter $\tau = 1$. (*Hint:* Use the h-method.)