## 3.8 Transformations

Let $X$ be a random variable, and let $g$ be some function. Because $X$ is random, the new variable $Y = g(X)$ is also random. Given what we know about $X$, what is the distribution of $Y$? This simple question will preoccupy us for several pages. Let's see some examples...

**Example 3.26.** Let $Y \sim \text{Unif}(-\pi/2, \pi/2)$, and define $X = \tan Y$. What is the distribution of $X$? Recall that the cdf of $Y$ is

$$F_Y(y) = \begin{cases} 0 & y \leq -\frac{\pi}{2} \\ \frac{y-\left(-\frac{\pi}{2}\right)}{\frac{\pi}{2}-\left(-\frac{\pi}{2}\right)} & -\frac{\pi}{2} < y < \frac{\pi}{2} \\ 1 & \frac{\pi}{2} \leq y. \end{cases}$$

So when $-\frac{\pi}{2} < y < \frac{\pi}{2}$,

$$F_Y(y) = \frac{y + \frac{\pi}{2}}{\pi} = \frac{y}{\pi} + \frac{1}{2}.$$

Anyhow, we want derive the pdf of $X = \tan Y$. For arbitrary $x \in \mathbb{R}$, we know that

$$\begin{aligned} F_X(x) &= P(X \leq x) \\ &= P(\tan Y \leq x) \\ &= P(Y \leq \tan^{-1} x) \\ &= F_Y(\tan^{-1} x) \\ &= \frac{\tan^{-1} x}{\pi} + \frac{1}{2}. \end{aligned}$$

So

$$f_X(x) = \frac{\mathrm{d}}{\mathrm{d}x} F_X(x) = \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\tan^{-1} x}{\pi} + \frac{1}{2}\right) = \frac{1}{\pi(1+x^2)}, \quad x \in \mathbb{R}.$$

**Definition 3.21.** An absolutely continuous random variable $X$ has the **Cauchy distribution** on $\text{Range}(X) = \mathbb{R}$ if its density is

$$f_X(x) = \frac{1}{\pi(1+x^2)}, \quad x \in \mathbb{R}. \tag{3.29}$$

The Cauchy density is plotted in Figure 3.30. There are two things to notice:

- both pdfs are symmetric and "centered" at zero;
- the Cauchy pdf has heavier tails than the standard normal pdf.

This second observation should make us worry about whether or not the Cauchy distribution has finite moments. It turns out that none of the raw moments of the Cauchy distribution either exist or are finite. Consider the mean. We say $E(X)$ exists if $E(|X|)$ is finite, so let's check it:

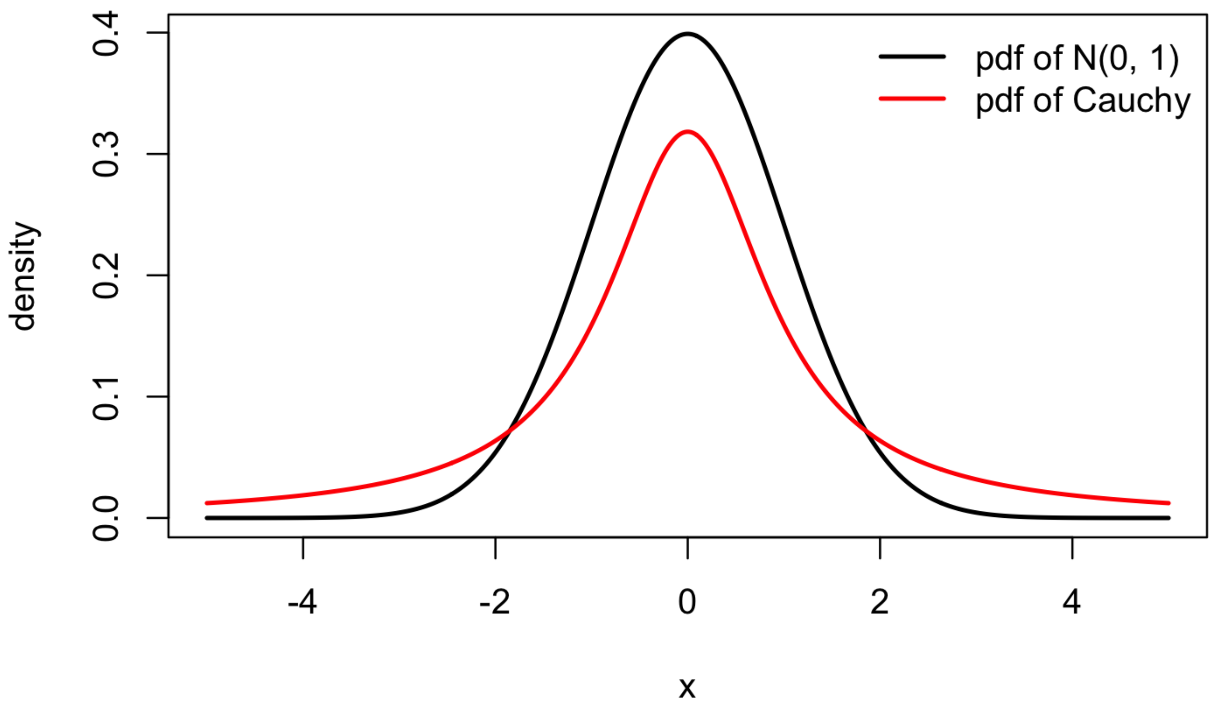


Figure 3.30: The Cauchy pdf compared to the standard normal. It has much heavier tails.

$$
\begin{aligned}
E(|X|) &= \int_{-\infty}^{\infty} |x| f(x)\,\mathrm{d}x \\
&= \frac{2}{\pi}\int_0^{\infty} \frac{x}{1+x^2}\,\mathrm{d}x \\
&= \lim_{t\to\infty} \frac{2}{\pi}\int_0^{t} \frac{x}{1+x^2}\,\mathrm{d}x \\
&= \lim_{t\to\infty} \frac{2}{\pi}\left[\frac{\ln(1+x^2)}{2}\right]_0^t \\
&= \lim_{t\to\infty} \frac{1}{\pi}[\ln(1+t^2) - \ln 1] \\
&= \lim_{t\to\infty} \frac{1}{\pi}\ln(1+t^2) \\
&= \infty.
\end{aligned}
$$

Voilà. Because the first moment does not exist, we don't have any of the higher moments either. So none of the raw moments are finite, and the Cauchy distribution does not have a moment-generating function.

This may irk you. You might be staring at the plot and thinking "what do you mean the expectation doesn't exist? There is clearly a center to this thing. Is that not the expected value?" No, it is not. As we have seen, the "center" of the distribution is not synonymous with the expected

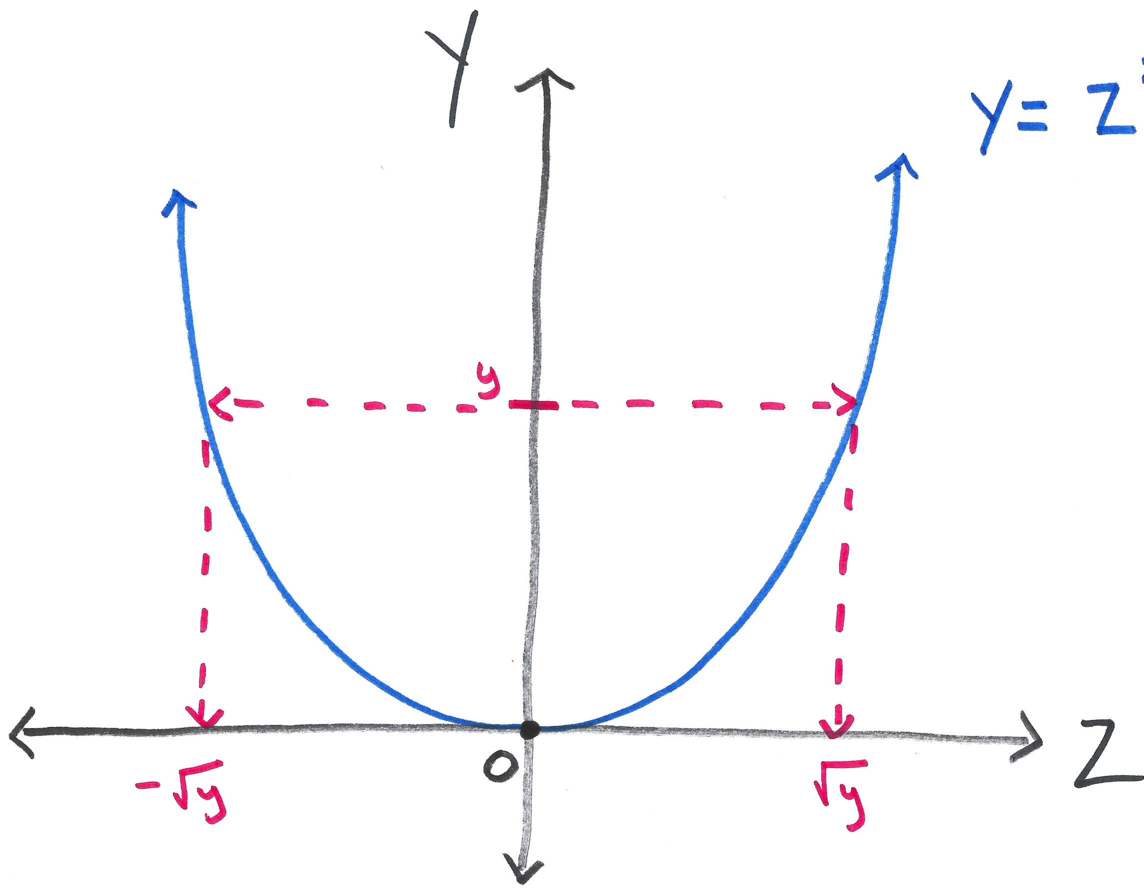


Figure 3.31: The event that $Z^2 \leq y$ is the same as $-\sqrt{y} \leq Z \leq \sqrt{y}$.

value. The expectation of a random variable is exactly defined to be the value of a particular integral $\int_{-\infty}^{\infty} x f(x)\mathrm{d}x$. If this integral diverges or does not exist, then the expected value is infinite or does not exist. I don't care where the "center" is. When you say "there is clearly a center to this thing," you are observing the *mode* of the pdf, and the median of the pdf, both of which you could compute for the Cauchy distribution (try it!):

$$0 = \underset{x\in\mathbb{R}}{\arg\max}\, f_X(x) \qquad \text{(mode)}$$
$$0 = F_X^{-1}(1/2) \qquad \text{(median)}$$

But the mean does not exist. Why not? Well, consider the definition of the expected value: $\int_{-\infty}^{\infty} x f(x)\mathrm{d}x$. The integrand is a product of two pieces: $x$ and $f(x)$. As $x \to \pm\infty$, the $x$ part blows up in magnitude, and the $f(x)$ part decays to 0. If $f(x)$ does not decay to 0 *fast enough* to swamp the growth of $x$, then the integral diverges. In the case of the normal distribution, the density decays fast enough for the expectation to exist. In the case of the Cauchy distribution, the decay is too slow.

**Example 3.27.** If $Z \sim \mathrm{N}(0,\ 1)$, then what is the density of $Y = Z^2$? Note that $\mathrm{Range}(Y) = (0,\ \infty)$, so fix arbitrary $y > 0$. Then

$$\begin{aligned} F_Y(y) &= P(Y \le y) \\ &= P\left(Z^2 \le y\right) \\ &= P\left(-\sqrt{y} \le Z \le \sqrt{y}\right) && \text{(see Figure 3.31)} \\ &= F_Z\left(\sqrt{y}\right) - F_Z\left(-\sqrt{y}\right). \end{aligned}$$

So

$$\begin{aligned} f_Y(y) &= \frac{\mathrm{d}}{\mathrm{d}y} F_Y(y) \\ &= \frac{\mathrm{d}}{\mathrm{d}y}\left[F_Z\left(\sqrt{y}\right) - F_Z\left(-\sqrt{y}\right)\right] \\ &= \frac{1}{2} y^{-1/2} f_Z\left(\sqrt{y}\right) + \frac{1}{2} y^{-1/2} f_Z\left(-\sqrt{y}\right) && F_Z'(z) = f_Z(z) \text{ and chain rule} \\ &= \frac{1}{2} y^{-1/2} f_Z\left(\sqrt{y}\right) + \frac{1}{2} y^{-1/2} f_Z\left(\sqrt{y}\right) && f_Z \text{ is symmetric about } 0 \\ &= y^{-1/2} f_Z\left(\sqrt{y}\right) && \text{combine like terms} \\ &= y^{-1/2} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\sqrt{y}\right)^2} \\ &= \frac{1}{2^{1/2}\sqrt{\pi}} y^{\frac{1}{2}-1} e^{-\frac{1}{2}y} \\ &= \frac{(1/2)^{1/2}}{\Gamma(1/2)} y^{\frac{1}{2}-1} e^{-\frac{1}{2}y}, \quad y > 0 && \Gamma(1/2) = \sqrt{\pi}. \end{aligned}$$

So $Y \sim \text{Gamma}(\alpha = 1/2,\ \beta = 1/2)$. Equivalently, we say $Y \sim \chi^2_1$.

**Theorem 3.17.** Let $X$ be absolutely continuous with pdf $f_X$ and cdf $F_X$, and let $a > 0$ and $b \in \mathbb{R}$ be constants. If we define a new random variable $Y = aX + b$ by applying a **location-scale transformation**, then

$$F_Y(y) = F_X\left(\frac{y-b}{a}\right). \tag{3.30}$$

$$f_Y(y) = \frac{1}{a} f_X\left(\frac{y-b}{a}\right). \tag{3.31}$$

*Proof.* For $y \in \mathbb{R}$, the cdf of $Y$ is

$$\begin{aligned} F_Y(y) &= P(Y \le y) \\ &= P(aX + b \le y) \\ &= P(aX \le y - b) \\ &= P\left(X \le \frac{y-b}{a}\right) \\ &= F_X\left(\frac{y-b}{a}\right). \end{aligned}$$

So by the chain rule, the density is

$$f_Y(y) = \frac{\mathrm{d}}{\mathrm{d}y} F_X\left(\frac{y-b}{a}\right) = \frac{\mathrm{d}}{\mathrm{d}y}\left(\frac{y-b}{a}\right) F'_X\left(\frac{y-b}{a}\right) = \frac{1}{a} f_X\left(\frac{y-b}{a}\right).$$

□

**Example 3.28.** We can use Theorem 3.17 to finally derive something basic but important: the mean and variance of the normal distribution. We know from Example 3.25 that if $Z \sim \mathrm{N}(0, 1)$, then $E(Z) = 0$ and $\mathrm{var}(Z) = 1$. But if $X \sim \mathrm{N}(\mu, \sigma^2)$, we have yet to actually derive that $E(X) = \mu$ and $\mathrm{var}(X) = \sigma^2$. Let $Z \sim \mathrm{N}(0, 1)$, fix $\sigma > 0$ and $\mu \in \mathbb{R}$, and let $X = \sigma Z + \mu$. From Theorem 3.17, we see that

$$\begin{aligned} f_X(x) &= \frac{1}{\sigma} f_Z\left(\frac{x-\mu}{\sigma}\right) \\ &= \frac{1}{\sigma}\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right). \end{aligned}$$

Comparing this to (3.20), we see that $X \sim \mathrm{N}(\mu, \sigma^2)$. But we also know from linearity that $E(X) = E(\sigma Z + \mu) = \sigma E(Z) + \mu = \sigma \cdot 0 + \mu = \mu$, and from Theorem 3.5 that $\mathrm{var}(X) = \mathrm{var}(\sigma Z + \mu) = \sigma^2 \mathrm{var}(Z) = \sigma^2 \cdot 1 = \sigma^2$. So there you have it: the general normal is just a location-scale transformation of the standard normal.

**Theorem 3.18.** (**Change of variables**) Let $X$ be an absolutely continuous random variable with pdf $f_X$, and let $g$ be a differentiable and strictly monotone function that is defined for all $x \in \text{Range}(X)$. Then the pdf of the new random variable $Y = g(X)$ is given by

$$f_Y(y) = f_X\left(g^{-1}(y)\right)\left|\frac{\mathrm{d}}{\mathrm{d}y}g^{-1}(y)\right|, \quad y \in \text{Range}(Y). \tag{3.32}$$

*Proof.* Because $g$ is differentiable and strictly monotone, $g^{-1}$ exists and is also differentiable and strictly monotone. We complete the proof by considering cases on $g$.

**Case 1: $g$ is strictly increasing**

If $g$ is strictly increasing, then $g^{-1}$ is strictly increasing also. This means that $\frac{\mathrm{d}}{\mathrm{d}y}g^{-1}(y)$ must be positive for all $y$. It also means that $g^{-1}$ is order-*preserving*. If $a \leq b$, then $g^{-1}(a) \leq g^{-1}(b)$.

Fix arbitrary $y \in \text{Range}(Y)$. Then:

$$\begin{aligned}
F_Y(y) &= P(Y \leq y) \\
&= P\left(g(X) \leq y\right) \\
&= P\left(X \leq g^{-1}(y)\right) \\
&= F_X\left(g^{-1}(y)\right).
\end{aligned}$$

Taking a derivative and applying the chain rule gives

$$\begin{aligned}
f_Y(y) &= \frac{\mathrm{d}}{\mathrm{d}y}F_X\left(g^{-1}(y)\right) \\
&= f_X\left(g^{-1}(y)\right)\underbrace{\frac{\mathrm{d}}{\mathrm{d}y}g^{-1}(y)}_{\text{positive!}} \\
&= f_X\left(g^{-1}(y)\right)\left|\frac{\mathrm{d}}{\mathrm{d}y}g^{-1}(y)\right|.
\end{aligned}$$

**Case 2: $g$ is strictly decreasing**

If $g$ is strictly decreasing, then $g^{-1}$ is strictly decreasing also. This means that $\frac{\mathrm{d}}{\mathrm{d}y}g^{-1}(y)$ must be negative for all $y$! It also means that $g^{-1}$ is order-*reversing*. If $a \leq b$, then $g^{-1}(a) \geq g^{-1}(b)$.

Fix arbitrary $y \in \text{Range}(Y)$. Then:

$$\begin{aligned}
F_Y(y) &= P(Y \leq y) \\
&= P\left(g(X) \leq y\right) \\
&= P\left(X > g^{-1}(y)\right) \\
&= 1 - P\left(X \leq g^{-1}(y)\right) \\
&= 1 - F_X\left(g^{-1}(y)\right).
\end{aligned}$$

Taking a derivative and applying the chain rule gives

$$\begin{aligned}
f_Y(y) &= \frac{\mathrm{d}}{\mathrm{d}y}[1 - F_X\left(g^{-1}(y)\right)] \\
&= -f_X\left(g^{-1}(y)\right)\underbrace{\frac{\mathrm{d}}{\mathrm{d}y}g^{-1}(y)}_{\text{negative!}} \\
&= f_X\left(g^{-1}(y)\right)\left|\frac{\mathrm{d}}{\mathrm{d}y}g^{-1}(y)\right|.
\end{aligned}$$

In either case, we got to the same place.

□

**Example 3.29.** If we redo Example 3.26 using the change of variables formula, we get the same result:

**Example 3.30.** If we redo Example 3.17 using the change of variables formula, we get the same result:

**Remark 3.20.** The function $g(x) = x^2$ is not invertible, so our change of variables formula could not, for example, be immediately applied to Example 3.27 where we showed that the square of a standard normal is chi-squared.

To answer the question "what is the distribution of this new random variable," we have so far answered it by figuring out what the cdf or the pdf is, precisely because we know that these objects uniquely characterize a distribution. The mgf is another such object, and so we can also determine the distribution of a transformation by deriving its mgf. Here is an example of how the mgf behaves under the location-scale transformation:

**Theorem 3.19.** Let $X$ be any random variable whose moment-generating function exists, and define a new random variable $Y = aX + b$ for arbitrary constants $a, b \in \mathbb{R}$. Then

$$M_Y(t) = e^{bt} M_X(at). \tag{3.33}$$

*Proof.*

$$\begin{aligned}
M_Y(t) &= M_{aX+b}(t) \\
&= E[e^{t(aX+b)}] \\
&= E[e^{taX+tb)}] \\
&= E[e^{taX} e^{tb}] \\
&= e^{tb} E[e^{taX}] \\
&= e^{tb} E[e^{(ta)X}] \\
&= e^{bt} M_X(at).
\end{aligned}$$

□

**Example 3.31.** If $Z \sim \mathrm{N}(0, 1)$, then $X = \sigma Z + \mu \sim \mathrm{N}(\mu, \sigma^2)$ as we saw in Example 3.28. We saw in Example 3.25 that $M_Z(t) = \exp(t^2/2)$, so we know from Theorem 3.19 that

$$M_X(t) = e^{\mu t} M_Z(\sigma t) = e^{\mu t} e^{(\sigma t)^2/2} = e^{\mu t + \frac{\sigma^2}{2} t^2}, \quad t \in \mathbb{R}.$$