

hw4-report

Jack McShane

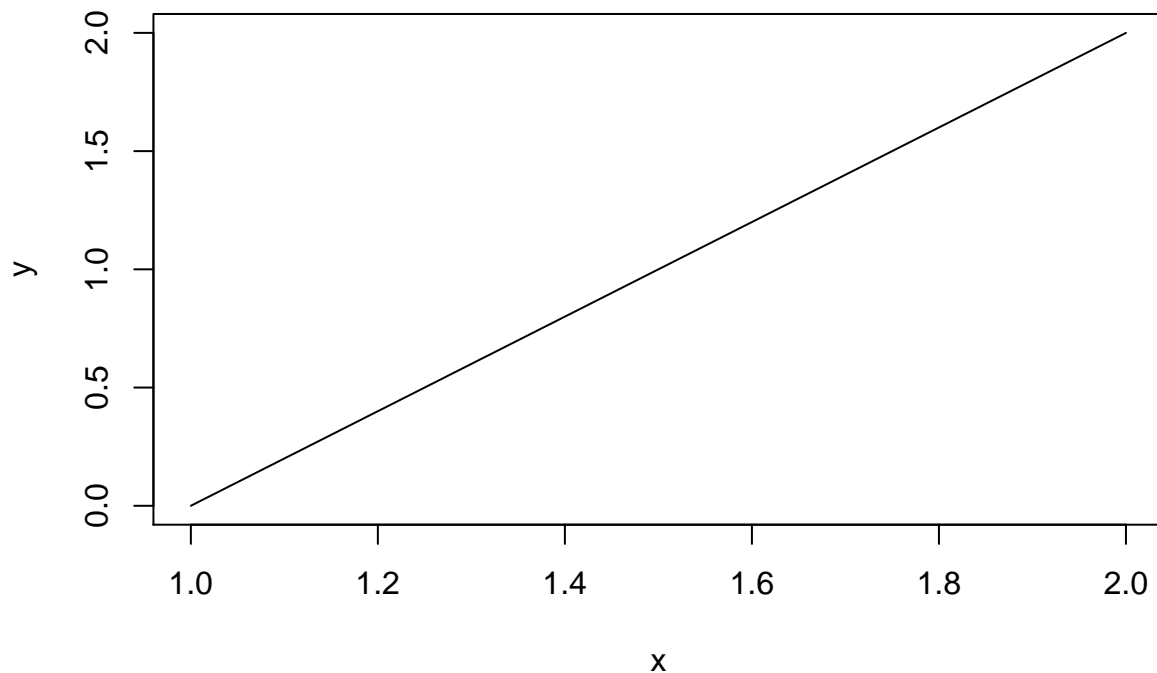
2022-03-01

1. Section 5.6 #2: Suppose that X is a continuous random variable with probability density function (pdf) f defined as follows:

$$f(x) = \begin{cases} 0 & x < 1 \\ 2(x-1) & 1 \leq x \leq 2 \\ 0 & x > 2 \end{cases}$$

a) Graph f .

```
fx = function(x){ (x-1)*2 }  
curve(fx, from=1, to=2, xlab="x", ylab="y")
```



b) Verify that f is a pdf.

- f is not a valid pdf, as the probability of any one value cannot exceed 1. At $x = 2$, the probability of the value is 2, and f is therefore not a valid pdf function.

2. Section 5.6 #3: Consider the function $f: \mathbb{R} \rightarrow \mathbb{R}$ defined by

$$f(x) = \begin{cases} 0 & x < 0 \\ cx & 0 < x < 1.5 \\ c(3-x) & 1.5 < x < 3 \\ 0 & x > 3 \end{cases}$$

where c is an undetermined constant.

a) For what value of c is f a probability density function?

- The value of c that makes this equation a valid probability function is $c = .44$. This can be determined using the property of random variables that the cumulative probability for all values of that variable must sum to exactly 1. Using this property, c can be calculated in the following manner:

$$\begin{aligned} \int f(x)dx &= 1 \\ \int_0^{1.5} cx dx + \int_{1.5}^3 c(3-x)dx &= 1 \\ \frac{c}{2}x^2 \Big|_0^{1.5} + c[3x - \frac{1}{2}x^2]_{1.5}^3 &= 1 \\ \frac{c}{2}(1.5^2 - 0^2) + c[(3(3) - \frac{1}{2}3^2) - (3(1.5) - \frac{1}{2}1.5^2)] &= 1 \\ \frac{2.25}{2}c + [(9 - \frac{1}{2}9) - (4.5 - \frac{1}{2}2.25)] &= 1 \\ (1.125 + 1.125)c &= 1 \\ c &= \frac{1}{2.25} = .44 \end{aligned}$$

3. Section 5.6 #6: Suppose that X has a normal distribution. Does $|X|$ have a normal distribution? Why or why not?

- $|X|$ does not follow a normal distribution. Given the possibility for X to take negative values, the absolute value operator would essentially flip the negative aspect of the associated distribution across the y-axis, forming a non-symmetrical distribution. It therefore follows that $|X|$ does not follow a normal distribution.

4. Section 5.6 #7: Let X be a normal random variable with mean $\mu = -5$ and standard deviation $\sigma = 10$. Compute the following and express the answers in the standard normal cdf $\Phi(\cdot)$ notation:

a) $P(X < 0)$

- Z transformation:

$$Z = \frac{x - \mu}{\sigma}$$

$$Z = \frac{0 - (-5)}{10}$$

$$Z = .5$$

- Solving:

$$P(X < 0) = P(Z < .5)$$

$$= \Phi(.5)$$

- This can be calculated in R using:

```
prob <- pnorm(.5)
```

```
## [1] "P(Z < .5) = 0.691462461274013"
```

b) $P(X > 5)$

- Z transformation:

$$Z = \frac{5 - (-5)}{10}$$

$$Z = 1$$

- Solving:

$$P(X > 5) = 1 - P(X < 5)$$

$$= 1 - P(Z < 1)$$

$$= 1 - \Phi(1)$$

- This can be calculated in R using:

```
prob <- pnorm(1)
```

```
## [1] "P(Z < 1) = 0.841344746068543"
```

c) $P(-3 < X < 7)$

- Z transformation

$$Z_1 = \frac{7 - (-5)}{10} \quad Z_2 = \frac{-3 - (-5)}{10}$$

$$Z_1 = 1.2 \quad Z_2 = .2$$

- Solving:

$$\begin{aligned}
P(-3 < X < 7) &= P(X < 7) - P(X > -3) \\
&= P(X < 7) - [1 - P(X < -3)] \\
&= P(Z < 1.2) - [1 - P(Z < .2)] \\
&= \Phi(1.2) - [1 - \Phi(.2)]
\end{aligned}$$

- This can be calculated in R using:

```
prob <- pnorm(1.2) - (1 - pnorm(.2))
```

```
## [1] "P(Z < 1.2) - [1 - P(Z < .2)] = 0.464190039217395"
```

d) $P(|X + 5| < 10)$

- expanding the absolute value operator:

$$\begin{aligned}
|X + 5| &< 10 \\
\rightarrow -10 &< X + 5 < 10 \\
\rightarrow -15 &< X < 5
\end{aligned}$$

- Z transformations:

$$\begin{aligned}
Z_1 &= \frac{-15 - (-5)}{10} & Z_2 &= \frac{-5 - (-5)}{10} \\
Z_1 &= -1 & Z_2 &= 0
\end{aligned}$$

- Solving:

$$\begin{aligned}
P(|X + 5| < 10) &= P(-15 < X < 5) \\
&= P(X < 5) - P(X > -15) \\
&= P(X < 5) - [1 - P(X < -15)] \\
&= P(Z < 0) - [1 - P(Z < -1)] \\
&= \Phi(0) - [1 - \Phi(-1)]
\end{aligned}$$

- This can be calculated in R using:

```
prob <- pnorm(0) - (1 - pnorm(-1))
```

```
## [1] "P(Z < 0) - [1 - P(Z < -1)] = -0.341344746068543"
```

5. Section 5.6 #8: Suppose that $X_1 \sim \text{Normal}(1, 9)$ and $X_2 \sim \text{Normal}(3, 16)$ are independent. Determine the mean and variance of each of the following normal random variables:

a) $X_1 + X_2$

$$\mu_{1,2} = \mu_1 + \mu_2 \quad \sigma_{1,2}^2 = \sigma_1^2 + \sigma_2^2$$

$$\mu_{1,2} = 1 + 3 \quad \sigma_{1,2}^2 = 9 + 16$$

$$\mu_{1,2} = 4 \quad \sigma_{1,2}^2 = 25$$

b) $-X_2$

$$\mu_{2'} = -\mu_2 \quad \sigma_{2'}^2 = \sigma_2^2$$

$$\mu_{2'} = -3 \quad \sigma_{2'}^2 = 16$$

c) $X_1 - X_2$

$$\mu_{1,2} = \mu_1 + (-1)\mu_2 \quad \sigma_{1,2}^2 = \sigma_1^2 + (-1)^2\sigma_2^2$$

$$\mu_{1,2} = 1 - 3 \quad \sigma_{1,2}^2 = 9 + 16$$

$$\mu_{1,2} = -2 \quad \sigma_{1,2}^2 = 25$$

d) $2X_1$

$$\mu_{1'} = 2\mu_1 \quad \sigma_{1'}^2 = (2^2)\sigma_1^2$$

$$\mu_{1'} = 2(1) \quad \sigma_{1'}^2 = 4(9)$$

$$\mu_{1'} = 2 \quad \sigma_{1'}^2 = 36$$

e) $2X_1 - 2X_2$

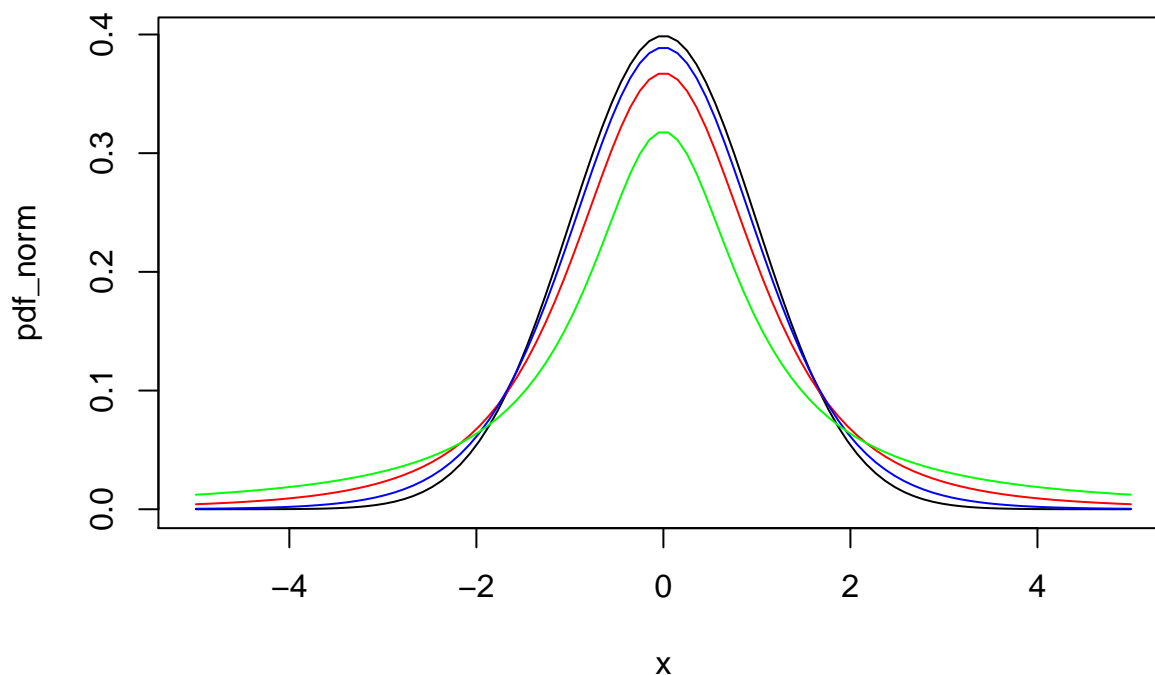
$$\mu_{1,2} = 2\mu_1 - 2\mu_2 \quad \sigma_{1,2}^2 = (2^2)\sigma_1^2 + (-2^2)\sigma_2^2$$

$$\mu_{1,2} = 2(1) - 2(3) \quad \sigma_{1,2}^2 = 4(9) + 4(16)$$

$$\mu_{1,2} = 2 - 6 \quad \sigma_{1,2}^2 = 36 + 64$$

$$\mu_{1,2} = -4 \quad \sigma_{1,2}^2 = 100$$

6. Draw the pdfs of $N(0, 1)$, $t(3)$, $t(10)$ and Cauchy distributions in one graph and comment on the figure.



- Much of the variability between the distributions comes from certainty of outcome. Normal distributions are generally meant to model population rather than some sample. If you have accurate data *for the population*, you can be rather certain of the probability for your r.v. to take a particular value. This explains the tall peak and the smaller tails. The rest of the distributions are used to model sampled data from the population, not the population itself. With this as the case, inherently, there is more uncertainty around the probability of any one outcome. This explains the lower peak around the mean and the larger tails at the ends.
- That said, the more degrees of freedom you have to model a problem, the more accurately you are able to model it. This explains the t10 distribution's better approximation of a normal distribution than the t3's.

7. Briefly describe the relationships among the following distributions: $N(\mu, \sigma^2)$, $N(0, 1)$, $\chi^2(n)$, $t(v)$, $F(v_1, v_2)$, **Cauchy**, and $U(0, 1)$.

- The $N(0,1)$ is simply a shifted, zero-centered form of the general normal distribution.
- The $\chi^2(n)$ distribution is simply the summation of various normal distributions raised to the 2nd power.
- The t distribution is a normal distribution over a $\chi^2(n)$ distribution (independent from the normal distribution in the numerator).

$$i.e. t_\mu = \frac{N(0, 1)}{\sqrt{\frac{\chi^2(v)}{v}}}$$

- $F(v_1, v_2)$ is simply the ratio of two $\chi^2(v)$ distributions over their respective degrees of freedom.

$$F(v_1, v_2) = \frac{\frac{\chi_1^2}{v_1}}{\frac{\chi_2^2}{v_2}}$$

- The Cauchy distribution is the ratio of two zero-centered normal distribution
- Each of these distributions are linked by the Transformation Integral Theory which describes the cumulative density functions of each of the aforementioned distributions as behaving like a uniform distribution.