Chapter 4. Continuous Random Variables

4.3: The Normal/Gaussian Random Variable

Slides (Google Drive) Alex Tsun Video (YouTube)

4.3.1 Standardizing RVs

Let's say you took two tests. You got 90% on history, and 50% on math. In which test did you do "better"? You might think it's obviously history, but actually your performance depends on the mean and standard deviation of scores in the class! We need to compare them on a fair playing ground then - this process is called standardizing. Let's see which test you truly did better on, given some extra information.

1. On your history test, you got a 90% when the mean was 70% and the standard deviation was 10%.
2. On your math test, you got a 50% when the mean was 35% and the standard deviation was 5%.

You scored higher in history, but how many standard deviations above the mean?

$$\frac{\text{your history score} - \text{mean history score}}{\text{standard deviation of history scores}} = \frac{90-70}{10} = 2$$

On your math test,

$$\frac{\text{your math score} - \text{mean math score}}{\text{standard deviation of math scores}} = \frac{50-35}{5} = 3$$

Then, in terms of standard deviations above the mean, you actually did better in math! What we just computed here was

$$\frac{X-\mu}{\sigma}$$

in order to calculate the number of standard deviations above the mean a random variable's value is. (Note how we are using standard deviation instead of variance here so the units are the same!)

Recall that in general, if X is any random variable (discrete or continuous) with $\mathbb{E}[X] = \mu$ and $\mathsf{Var}(X) = \sigma^2$, and $a,\ b \in \mathbb{R}$. Then,

$$\mathbb{E}[aX+b] = a\mathbb{E}[X] + b = a\mu + b$$
$$\mathsf{Var}(aX+b) = a^2\mathsf{Var}(X) = a^2\sigma^2$$

In particular, we call $\frac{X-\mu}{\sigma}$ a standardized version of X, as it measures how many standard deviations above the mean a point is. We standardize random variables for fair comparison. Applying linearity of expectation and variance of random variables to standardized random variables, we get the expectation and variance of standardized random variables:

$$\mathbb{E}\left[\frac{X-\mu}{\sigma}\right] = \frac{1}{\sigma}(\mathbb{E}[X] - \mu) = 0$$

$$\mathsf{Var}\left(\frac{X-\mu}{\sigma}\right) = \frac{1}{\sigma^2}\mathsf{Var}(X-\mu) = \frac{1}{\sigma^2}\mathsf{Var}(X) = \frac{1}{\sigma^2}\sigma^2 = 1 \implies \sigma_X = \sqrt{\mathsf{Var}(X)} = \sqrt{1} = 1$$

It turns out the mean is 0 and the standard deviation (and variance) is 1 ! This makes sense because on average, someone is average (0 standard deviations above the mean), and the standard deviation is 1.

4.3.2 The Normal/Gaussian Random Variable

Definition 4.3.1: Normal (Gaussian, "bell curve") distribution

$X \sim \mathcal{N}(\mu, \sigma^2)$ if and only if X has the following PDF (and range $\Omega_X = (-\infty, +\infty)$):

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

where $\exp(y) = e^y$. This Normal random variable actually has as parameters its mean and variance, and hence:

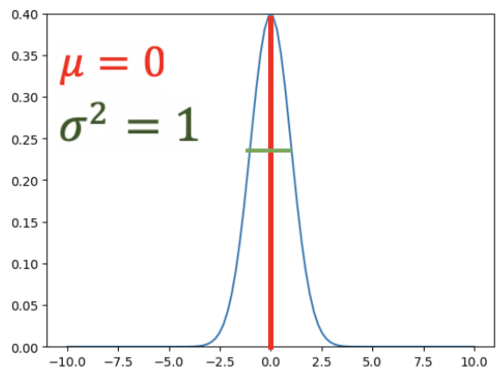
$$\mathbb{E}[X] = \mu$$

$$\mathsf{Var}(X) = \sigma^2$$

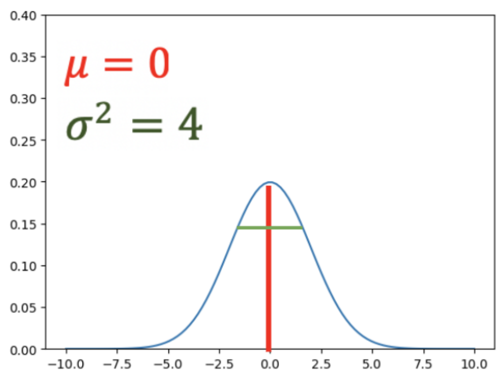
Unfortunately, there is no closed form formula for the CDF (there wasn't one for the Gamma RV) either. We'll see how to compute these probabilities anyway though soon using a lookup table!

Normal distributions produce bell-shaped curves. Here are some visualizations of the density function for varying μ and σ^2.

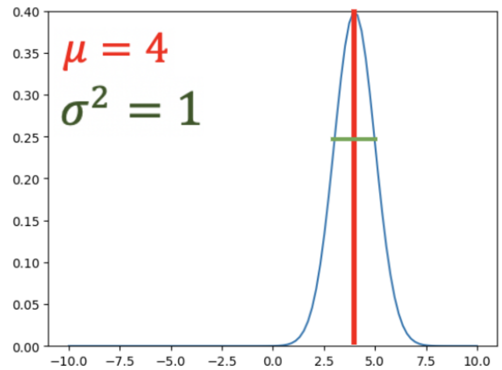
For instance, a normal distribution with $\mu = 0$ and $\sigma = 1$ produces the following bell curve:



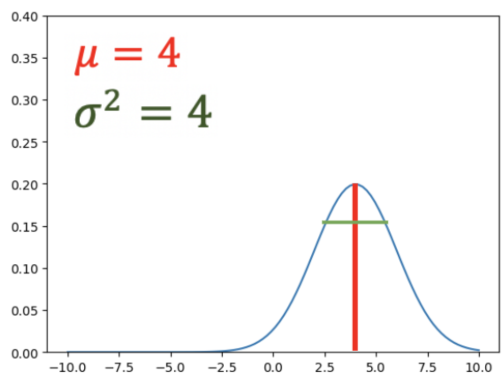
If the standard deviation increases, it becomes more likely for the variable to be farther away from the mean, so the distribution becomes flatter. For instance, a curve with the same $\mu = 0$ but higher $\sigma = 2$ ($\sigma^2 = 4$) looks like this:



If you change the mean, the distribution will shift left or right. For instance, increasing the mean so $\mu = 4$ shifts the distribution 4 to the right. The shape of the curve remains unchanged:



If you change the mean AND standard deviation, the curves shape changes and shifts. For instance, changing the mean so $\mu = 4$ and standard deviation so $\sigma = 2$ gives us a flatter, shifted curve:



4.3.3 Closure Properties of the Normal Random Variable

Occasionally, when we sum two independent random variables of the same type, we get the same type. For example, if $X \sim \text{Bin}(n, p)$ and $Y \sim \text{Bin}(m, p)$ are independent, then $X + Y \sim \text{Bin}(n + m, p)$ because X is the number of successes in n trials, and Y is the number of successes in m trials. It also turns out similar properties hold for the Poisson, Negative Binomial, and Gamma random variables when you think of their English meaning. We'll formally prove some of these results in 5.5 though.

However, scaling and shifting a random variable often does not keep it in the same family. Continuous uniform rvs are the only ones we learned so far that do: if $X \sim \text{Unif}(0, 1)$, then $3X + 2 \sim \text{Unif}(2, 5)$: we'll learn how to prove this in the next section! However, this is not true for the others; for example, the range of a $\text{Poi}(\lambda)$ is $\{0, 1, 2, \ldots\}$ as it is the number of events in a unit of time, and $2X$ has range $\{0, 2, 4, 6, \ldots, \}$ so cannot be Poisson (cannot be an odd number)! We'll see that Normal random variables have these closure properties.

Recall that in general, if X is any random variable (discrete or continuous) with $\mathbb{E}[X] = \mu$ and $\mathsf{Var}(X) = \sigma^2$, and $a,\ b \in \mathbb{R}$. Then,

$$\mathbb{E}\left[aX+b\right] = a\mathbb{E}\left[X\right] + b = a\mu + b$$
$$\mathsf{Var}\left(aX+b\right) = a^2\mathsf{Var}\left(X\right) = a^2\sigma^2$$

Definition 4.3.2: Closure of the Normal Under Scale and Shift

If $X \sim \mathcal{N}(\mu, \sigma^2)$, then $aX + b \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$.
In particular,

$$\frac{X-\mu}{\sigma} \sim \mathcal{N}(0,1)$$

We will prove this theorem later in section 5.6 using Moment Generating Functions! This is really amazing - the mean and variance are no surprise. The fact that scaling and shifting a Normal random variable results in another Normal random variable is very interesting!

Let X, Y be ANY *independent* random variables (discrete or continuous) with $\mathbb{E}\left[X\right] = \mu_X$, $\mathbb{E}\left[Y\right] = \mu_Y$, $\mathsf{Var}\left(X\right) = \sigma_X^2$, $\mathsf{Var}\left(Y\right) = \sigma_Y^2$ and $a, b, c \in \mathbb{R}$. Recall,

$$\mathbb{E}\left[aX+bY+c\right] = a\mathbb{E}\left[X\right] + b\mathbb{E}\left[Y\right] + c = a\mu_X + b\mu_Y + c$$
$$\mathsf{Var}\left(aX+bY+c\right) = a^2\mathsf{Var}\left(X\right) + b^2\mathsf{Var}\left(Y\right) = a^2\sigma_X^2 + b^2\sigma_Y^2$$

Definition 4.3.3: Closure of the Normal Under Addition

If $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$ and $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$ (both independent normal random variables), then

$$aX + bY + c \sim \mathcal{N}(a\mu_X + b\mu_Y + c, a^2\sigma_X^2 + b^2\sigma_Y^2)$$

Again, this is really amazing. The mean and variance aren't a surprise again, but the fact that adding two independent Normals results in another Normal distribution is not trivial, and we will prove this later as well!

Example(s)

Suppose you believe temperatures in the Vancouver, Canada each day are approximately normally distributed with mean 25 degrees Celsius and standard deviation 5 degrees Celsius. However, your American friend only understands Fahrenheit.

1. What is the distribution of temperatures each day in Vancouver in Fahrenheit? To convert Celsius (C) to Fahrenheit (F), the formula is $F = \frac{9}{5}C + 32$.

2. What is the distribution of the average temperature over a week in the Vancouver, in Fahrenheit? That is, if you were to sample a random week's average temperature, what is its distribution? Assume the temperature each day is independent of the rest (this may not be a realistic assumption).

Solution

1. The degrees in Celsius are $\mathcal{N}(\mu_C = 25, \sigma_C^2 = 5^2)$. Since $F = \frac{9}{5}C + 32$, we know by linearity of expectation and properties of variance:

$$\mu_F = \mathbb{E}\left[F\right] = \mathbb{E}\left[\frac{9}{5}C + 32\right] = \frac{9}{5}\mathbb{E}\left[C\right] + 32 = \frac{9}{5}25 + 32 = 77$$

$$\sigma_F^2 = \mathsf{Var}\left(F\right) = \mathsf{Var}\left(\frac{9}{5}C + 32\right) = \left(\frac{9}{5}\right)^2 \mathsf{Var}\left(C\right) = \left(\frac{9}{5}\right)^2 5^2 = 81$$

These values are no surprise, but by closure of the Normal distribution, we can say that $F \sim \mathcal{N}(\mu_F = 77, \sigma_F^2 = 9^2)$.

2. Let $F_1, F_2, \ldots, F_7$ be independent temperatures over a week, so each $F_i \sim \mathcal{N}(\mu_F = 77, \sigma_F^2 = 81)$. Let $\bar{F} = \frac{1}{7}\sum_{i=1}^{7} F_i$ denote the average temperature over this week. Then, by linearity of expectation and properties of variance (requiring independence),

$$\mathbb{E}\left[\frac{1}{7}\sum_{i=1}^{7} F_i\right] = \frac{1}{7}\sum_{i=1}^{7}\mathbb{E}\left[F_i\right] = \frac{1}{7}\cdot 7 \cdot 77 = 77$$

$$\mathsf{Var}\left(\frac{1}{7}\sum_{i=1}^{7} F_i\right) = \frac{1}{7^2}\sum_{i=1}^{7}\mathsf{Var}\left(F_i\right) = \frac{1}{7^2}\cdot 7 \cdot 81 = \frac{81}{7}$$

Note that the mean is the same, but the variance is smaller. This might make sense because we expect the average temperature over a week should match that of a single day, but it is more stable (has lower variance). By closure properties of the Normal distribution, since we take a sum of independent Normal RVs and then divide it by 7, $\bar{F} = \frac{1}{7}\sum_{i=1}^{7} F_i \sim \mathcal{N}(\mu = 77, \sigma^2 = 81/7)$.

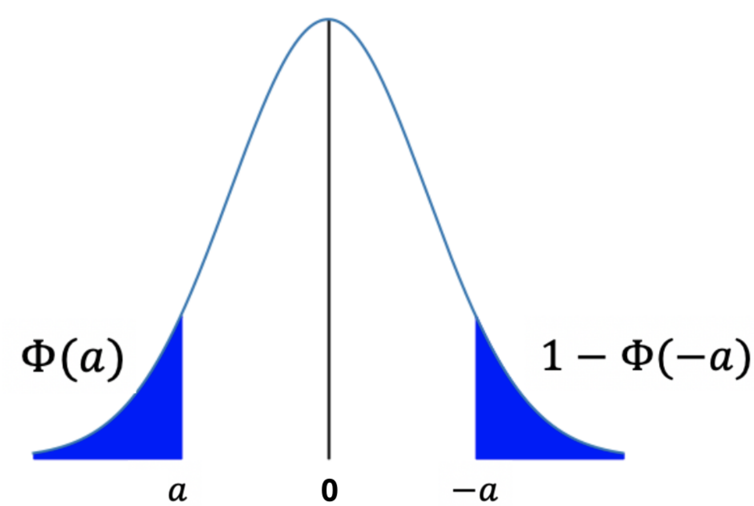
□

4.3.4 The Standard Normal CDF

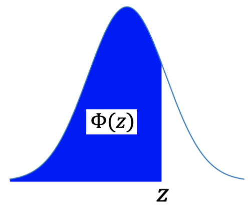
We'll finally learn how to use to calculate probabilities like $\mathbb{P}\left(X \leq 55\right)$ if X has a Normal distribution!

If $Z \sim \mathcal{N}(0, 1)$ is the **standard normal** (the normal RV with mean 0 and variance/standard deviation 1), we denote the CDF $\Phi(a) = F_Z(a) = \mathbb{P}\left(Z \leq a\right)$, since it is so commonly used. There is no closed-form formula, so this CDF is stored in a Φ table (read a "Phi Table"). Remember, $\Phi(a)$ is just the area to the left of a.

Since the normal distribution curve is symmetric, the area to the left of a is the same as the area to the right of $-a$. This picture below shows that $\Phi(a) = 1 - \Phi(-a)$.



To get the CDF $\Phi(1.09) = \mathbb{P}(Z \leq 1.09)$ from the Φ table, we look at the row with a value of 1.0, and column with value 0.09, as marked here:



Φ Table: $\mathbb{P}(Z \leq z)$ when $Z \sim \mathcal{N}(0, 1)$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.5279	0.53188	0.53586
0.1	0.53983	0.5438	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.6293	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.6591	0.66276	0.6664	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.7054	0.70884	0.71226	0.71566	0.71904	0.7224
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.7549
0.7	0.75804	0.76115	0.76424	0.7673	0.77035	0.77337	0.77637	0.77935	0.7823	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1.0	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.8665	0.86864	0.87076	0.87286	0.87493	0.87698	0.879	0.881	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
1.3	0.9032	0.9049	0.90658	0.90824	0.90988	0.91149	0.91309	0.91466	0.91621	**0.91774**
1.4	0.91924	0.92073	0.9222	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.9452	0.9463	0.94738	0.94845	0.9495	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.9608	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.9732	0.97381	0.97441	0.975	0.97558	0.97615	0.9767
2.0	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.9803	0.98077	0.98124	0.98169
2.1	0.98214	0.98257	0.983	0.98341	0.98382	0.98422	0.98461	0.985	0.98537	0.98574
2.2	0.9861	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.9884	0.9887	0.98899
2.3	0.98928	0.98956	0.98983	0.9901	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158
2.4	0.9918	0.99202	0.99224	0.99245	0.99266	0.99286	0.99305	0.99324	0.99343	0.99361
2.5	0.99379	0.99396	0.99413	0.9943	0.99446	0.99461	0.99477	0.99492	0.99506	0.9952
2.6	0.99534	0.99547	0.9956	0.99573	0.99585	0.99598	0.99609	0.99621	0.99632	0.99643
2.7	0.99653	0.99664	0.99674	0.99683	0.99693	0.99702	0.99711	0.9972	0.99728	0.99736
2.8	0.99744	0.99752	0.9976	0.99767	0.99774	0.99781	0.99788	0.99795	0.99801	0.99807
2.9	0.99813	0.99819	0.99825	0.99831	0.99836	0.99841	0.99846	0.99851	0.99856	0.99861
3.0	0.99865	0.99869	0.99874	0.99878	0.99882	0.99886	0.99889	0.99893	0.99896	0.999

From this, we see that $\mathbb{P}(Z \leq 1.39) = \Phi(1.39) \approx 0.91774$. (Look at the gray row 1.3, and the column 0.09).

This table usually only has positive numbers, so if you want to look up negative numbers, it's necessary to use the fact that $\Phi(a) = 1 - \Phi(-a)$. For example, if we want $\mathbb{P}(Z \leq -2.13) = \Phi(-2.13)$, we need to do $1 - \Phi(2.13) = 1 - 0.9834 = 0.0166$ (try to find $\Phi(2.13)$ yourself above).

How does this help though when X is Normal but not the standard normal? In general, for a $X \sim \mathcal{N}(\mu, \sigma^2)$, we can calculate the CDF of X by standardizing it to be standard normal,

$$
\begin{aligned}
F_X(y) &= \mathbb{P}(X \le y) && \text{[def of CDF]} \\
&= \mathbb{P}\left(\frac{X-\mu}{\sigma} \le \frac{y-\mu}{\sigma}\right) && \text{[standardizing both sides]} \\
&= \mathbb{P}\left(Z \le \frac{y-\mu}{\sigma}\right) && \left[\text{since } Z = \frac{X-\mu}{\sigma} \text{ is the standardized normal, } Z \sim \mathcal{N}(0,1)\right] \\
&= \Phi\left(\frac{y-\mu}{\sigma}\right) && \text{[def of } \Phi]
\end{aligned}
$$

We can also find $\mathbb{P}(a \le X \le b)$,

$$
\begin{aligned}
\mathbb{P}(a \le X \le b) &= F_X(b) - F_X(a) && \text{[def of CDF]} \\
&= \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right) && \text{[def of } \Phi]
\end{aligned}
$$

Definition 4.3.4: Standard Normal Random Variable

The "standard normal" random variable is typically denoted Z and has mean 0 and variance 1. By the closure property of normals, if $X \sim \mathcal{N}(\mu, \sigma^2)$, then $Z = \frac{X-\mu}{\sigma} \sim \mathcal{N}(0,1)$. The CDF has no closed form, but we denote the CDF of the standard normal by $\Phi(a) = F_Z(a) = \mathbb{P}(Z \le a)$. Note that by symmetry of the density about 0, $\Phi(-a) = 1 - \Phi(a)$.

See some examples below of how we can use the Φ table to calculate probabilities associated with the Normal distributions! Again, the Φ table gives the CDF of the standard Normal since it doesn't have a closed form like Uniform/Exponential. Also, *any* Normal RV can be standardized so we can look up probabilities in the Φ table!

Example(s)

Suppose the age of a random adult in the United States is (approximately) normally distributed with mean 50 and standard deviation 15.

1. What is the probability that a randomly selected adult in the US is over 70 years old?
2. What is the probability that a randomly selected adult in the US is under 25 years old?
3. What is the probability that a randomly selected adult in the US is between 40 and 45 years old?

Solution

1. The height of a random adult is $X \sim \mathcal{N}(\mu = 50, \sigma^2 = 15^2)$, so remember we standardize to use the

standard Gaussian:

$$
\begin{aligned}
\mathbb{P}(X > 70) &= \mathbb{P}\left(\frac{X-50}{15} > \frac{70-50}{15}\right) && \text{[standardize]} \\
&= \mathbb{P}(Z > 1.33) && \left[\frac{X-\mu}{\sigma} = Z \sim \mathcal{N}(0,1)\right] \\
&= 1 - \mathbb{P}(Z \le 1.33) && \text{[complement]} \\
&= 1 - \Phi(1.33) && \text{[def of } \Phi] \\
&= 1 - 0.9082 && \text{[look up } \Phi \text{ table from earlier]} \\
&= 0.0918
\end{aligned}
$$

2. We do a similar calculation:

$$
\begin{aligned}
\mathbb{P}(X < 25) &= \mathbb{P}\left(\frac{X-50}{15} > \frac{25-50}{15}\right) && \text{[standardize]} \\
&= \mathbb{P}(Z < -2/3) && \left[\frac{X-\mu}{\sigma} = Z \sim \mathcal{N}(0,1)\right] \\
&= \Phi(-0.67) && \text{[recall since continuous rv, identical to less than or equal]} \\
&= 1 - \Phi(0.67) && \text{[symmetry trick to make positive]} \\
&= 1 - 0.7486 && \text{[look up } \Phi \text{ table from earlier]} \\
&= 0.2514
\end{aligned}
$$

3. We do a similar calculation:

$$
\begin{aligned}
\mathbb{P}(40 < X < 45) &= \mathbb{P}\left(\frac{40-50}{15} < \frac{X-50}{15} < \frac{45-50}{15}\right) && \text{[standardize]} \\
&= \mathbb{P}(-2/3 < Z < -1/3) && \left[\frac{X-\mu}{\sigma} = Z \sim \mathcal{N}(0,1)\right] \\
&= \Phi(-0.33) - \Phi(-0.67) && [\mathbb{P}(a < X < b) = F_X(b) - F_X(a)] \\
&= (1 - \Phi(0.33)) - (1 - \Phi(0.67)) && \text{[symmetry trick to make positive]} \\
&= \Phi(0.67) - \Phi(0.33) \\
&= 0.7486 - 0.6293 && \text{[look up } \Phi \text{ table from earlier]} \\
&= 0.1193
\end{aligned}
$$

□

4.3.5 Exercises

1. Suppose the time (in hours) it takes for you to finish pset i is approximately $X_i \sim \mathcal{N}(\mu = 10, \sigma^2 = 9)$ (for $i = 1, \ldots, 5$) and the time (in hours) it takes for you to finish a project is approximately $Y \sim \mathcal{N}(\mu = 20, \sigma^2 = 10)$. Let $W = X_1 + X_2 + X_3 + X_4 + X_5 + Y$ be the time it takes to complete all 5 psets and the project.
   (a) What are the mean and variance of W?
   (b) What is the distribution of W and what are its parameter(s)?
   (c) What is the probability that you complete all the homework in under 60 hours?

Solution:

(a) The mean by linearity of expectation is $\mathbb{E}[W] = \mathbb{E}[X_1] + \cdots + \mathbb{E}[X_5] + \mathbb{E}[Y] = 50 + 20 = 70$. Variance adds for independent RVs, so $\mathsf{Var}(W) = \mathsf{Var}(X_1) + \cdots + \mathsf{Var}(X_5) + \mathsf{Var}(Y) = 45 + 10 = 55$.

(b) Since W is the sum of independent Normal random variables, W is also normal with the parameters we calculated above. So $W \sim \mathcal{N}(\mu = 70, \sigma^2 = 55)$.

(c)

$$\mathbb{P}(W < 60) = \mathbb{P}\left(\frac{W - 70}{\sqrt{55}} < \frac{60 - 70}{\sqrt{55}}\right) \approx \mathbb{P}(Z < -1.35) = \Phi(-1.35) = 1 - \Phi(1.35) = 1 - 0.9115 = 0.0885$$

.