Section 3.4. The Normal Distribution

Note. You see the importance of the normal distribution in, even, Introduction to Probability and Statistics (MATH 1530); see my online notes for this class on Chapter 3. The Normal Distributions and Chapter 11. Sampling Distributions (this second source includes a statement of the Central Limit Theorem, which we will see in Section 5.3 of these notes). In this section, we define the normal distribution and the standard normal distribution. We show some properties of these distributions.

Note. Consider the integral $\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\,dz$ (we'll see that the integrand is the standard normal distribution). First, we apply the Direct Comparison Test for improper integrals of nonnegative functions to show that this integral exists (see Theorem 2 of my online Calculus 2 [MATH 1920] notes on Section 8.7. Improper Integrals). Since the exponential function $\exp(x)$ is an increasing function then $\exp(-z^2/2) < \exp(-z+1)$ for all $z \geq 0$ since $-z^2/2 < -z+1$ for $z \geq 0$. Similarly $-z^2/2 \leq z+1$ for $z \leq 0$ and so $-z^2/2 \leq -|z|+1$ for all $z \in \mathbb{R}$. Since

$$\int_{-\infty}^{\infty} e^{-|z|+1}\,dz = 2\int_0^{\infty} e^{-z+1}\,dz = 2(-e^{-z+1})_0^{\infty}$$

$$= -2\lim_{b\to\infty}(e^{-z+1})|_0^b = -2\left(\lim_{b\to\infty} e^{-b+1}\right) + 2e = 0 + 2e = 2e,$$

then by the Direct Comparison Test, $\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\,dz$ exists.

Note. We can in fact show that for $I = \int_0^{\infty} \frac{2}{\sqrt{\pi}} e^{-x^2}\,dx$, we have

$$I^2 = \left(\int_0^{\infty} \frac{2}{\sqrt{\pi}} e^{-x^2}\,dx\right)^2 = \frac{4}{\pi}\left(\int_0^{\infty} e^{-x^2}\,dx\right)\left(\int_0^{\infty} e^{-y^2}\,dy\right)$$

$$= \frac{4}{\pi}\int_0^\infty \int_0^\infty e^{-(x^2+y^2)}\,dx\,dy = 1$$

with substitutions $x = r\cos\theta$, $y = r\sin\theta$, $r^2 = x^2+y^2$, and $dx\,dy = r\,dr\,d\theta$ (i.e., the use of polar coordinates; see Exercise 15.4.41 of my online Calculus 3 [MATH 2110] on Section 15.4. Double Integrals in Polar Form). From this, we have with the substitution $z = \sqrt{2}x$ that

$$1 = \int_0^\infty \frac{2}{\sqrt{\pi}} e^{-x^2}\,dx = \int_0^\infty \frac{2}{\sqrt{\pi}} e^{-z^2/2}\frac{1}{\sqrt{2}}\,dz = \int_0^\infty \sqrt{\frac{2}{\pi}} e^{-z^2/2}\,dz \qquad (*)$$

and so, by $(*)$,

$$\int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\,dz = 2\int_0^\infty \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\,dz = \int_0^\infty \sqrt{\frac{2}{\pi}} e^{-z^2/2}\,dz = 1.$$

Since $f(z) = \dfrac{1}{\sqrt{2\pi}} e^{-z^2/2}$ is nonnegative on $\mathbb{R}$ and integrates to 1, then f is a probability density function of a random variable, which we denote Z.

Note 3.4.A. By Definition 1.9.3, the moment generating function of Z is (for $t \in \mathbb{R}$):

$$M_Z(t) = E[e^{tZ}] = \int_{-\infty}^\infty e^{tz} f(z)\,dz = \int_{-\infty}^\infty e^{tz}\frac{1}{\sqrt{2\pi}} e^{-z^2/2}\,dz = \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}} e^{-z^2/2+tz}\,dz$$

$$= \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}} e^{-z^2/2+2tz/2-t^2/2+t^2/2}\,dt = e^{t^2/2}\int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}} e^{-(z^2-2tz+t^2)/2}\,dz$$

$$= e^{t^2/2}\int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}} e^{-(z-t)^2}\,dz = e^{t^2/2}(1) = e^{t^2/2}.$$

Note. Since we have the moment generating function $M_Z(t) = e^{t^2/2}$, then

$$M_Z'(t) = te^{t^2/2} \text{ and } M_Z''(t) = e^{t^2/2} + t^2 e^{t^2/2}.$$

By Note 1.9.B,

$$\mu = E(Z) = M_Z'(0) = 0 \text{ and } E[Z^2] = M_Z''(0) = 1.$$

So the variance of Z is $\text{Var}(Z) = \sigma^2 = M_Z''(0) - (M_Z'(0))^2 = (1) - (0)^2 = 1$.

Note. Define another (continuous) random variable X as $X = g(Z) = bZ + a$ where $b > 0$. Then $Z = g^{-1}(X) = (X - a)/b$ and so, by "Theorem 1.7.1. The Cumulative Distribution Function Technique," the probability density function of X is

$$f_X(x) = f_Z(g^{-1}(x))|dz/dx| = f_Z((x-a)/b)|1/b| = \frac{1}{\sqrt{2\pi}b}e^{-(x-a)^2/b^2} \text{ for } x \in \mathbb{R}.$$

Since $E(Z) = 0$ then $E((X - z)/b) = 0$ and so $E(X) = a$ (because expectation is linear by Theorem 1.8.2). By Theorem 1.9.1 (which implies $\text{Var}(aX + b) = a^2\text{Var}(X)$) we have

$$\text{Var}(Z) = b^2\text{Var}\left(\frac{1}{b}X - \frac{a}{b}\right) = \frac{1}{b^2}\text{Var}(X),$$

or $\text{Var}(X) = b^2\text{Var}(Z) = b^2(1) = b^2$. We now replace a with μ and b with σ to get the normal distribution.

Definition 3.4.1. A random variable X has a *normal distribution* if its probability density function is

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) \text{ for } z \in \mathbb{R}.$$

The parameters μ and σ^2 are the mean and variance of X, respectively. We denote this as the $N(\mu, \sigma^2)$ distribution. The distribution $N(0, 1)$ is the *standard normal distribution.*

Note 3.4.B. We can use the relationship and linearity of expectation (Theorem 1.8.2) to find the expectation of normally distributed random variable X:

$$\begin{aligned} M_X(t) &= E(e^{tx}) = E(t(\sigma Z + \mu)) = E(e^{\mu t}e^{t\sigma Z}) = e^{\mu t}E(e^{(t\sigma Z)}) \\ &= e^{\mu t}e^{(t\sigma)^2/2} \text{ since } E(e^{tZ}) = e^{t^2/2} \\ &= e^{\mu t + t^2\sigma^2/2} \text{ for } t \in \mathbb{R}. \end{aligned}$$

Of course if $\mu = 0$ and $\sigma = 1$, then this reduces to $M_Z(t)$.

Note. The graph of the probability density function of normally distributed random variable X is given in Figure 3.4.1. Notice that the graph is (1) symmetric with respect to the vertical line $x = \mu$, (2) has a maximum at $x = \mu$ of $1/(\sigma\sqrt{2\pi})$, (3) has a horizontal asymptote of $y = 0$, and (4) has points of inflection at $x = \mu \pm \sigma$ (as is to be verified in Exercise 3.4.7).

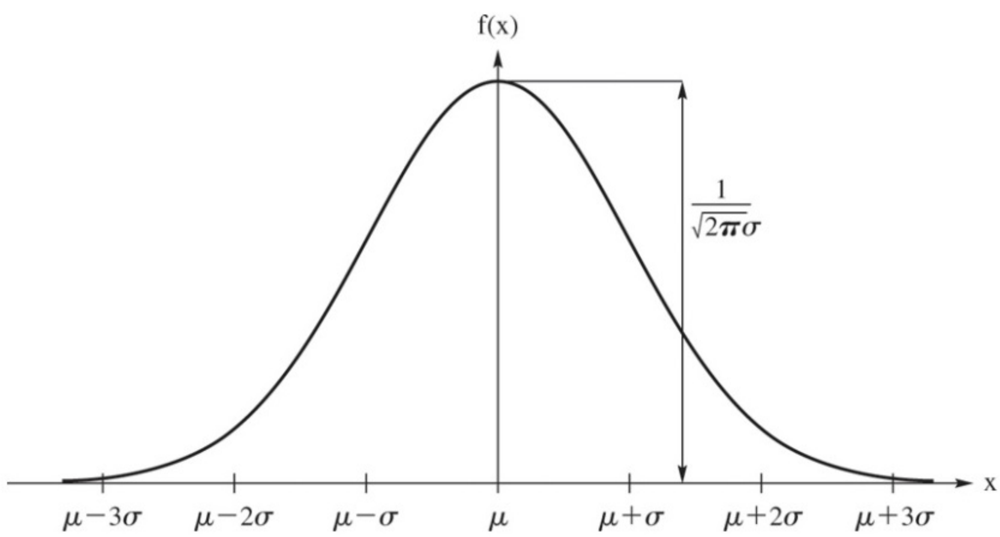

Figure 3.4.1. The normal distribution $N(\mu, \sigma^2)$.

Note. As you likely know, many statistical applications involve the cumulative area under the probability density function of the normal distribution:

$$P(X \leq x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma} e^{-(t-\mu)^2/(2\sigma^2)}\, dt.$$

Hogg, McKean, and Craig comment: "From calculus we know that the integrand [here] does not have an antiderivative…" (see page 189). This is not the case! There is not an antiderivative that can be expressed in terms of "elementary" functions. This is explained in some detail in Maxwell Rosenlicht's "Integration in Finite Terms," *American Mathematical Monthly*, **79**(0), 963–972 (1972) (in the first paragraph of this work, it is claimed that Joseph Liouville [1809-1882] proved that $\int e^{x^2}\, dx$ cannot be expressed in terms of elementary functions). However, we can find an antiderivative using series, Recall that $e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$ for $x \in \mathbb{R}$. So

$$e^{-x^2/2} = \sum_{k=0}^{\infty} \frac{(-x^2/2)^k}{k!} = \sum_{k=0}^{\infty} \frac{(-1)^k x^{2k}}{2^k k!} \text{ for } x \in \mathbb{R}.$$

Therefore, the indefinite integral of the probability density function of the standard normal distribution is

$$\int \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx = \frac{1}{\sqrt{2\pi}} \int \left(\sum_{k=0}^{\infty} \frac{(-1)^k x^{2k}}{2^k k!} \right) dx$$

$$= \frac{1}{\sqrt{2\pi}} \sum_{k=0}^{\infty} \frac{(-1)^k x^{2k+1}}{2^k (2k+1) k!} + C \text{ for } x \in \mathbb{R}.$$

So we have an explicit antiderivative, it is just given by a series. Notice that we have used the (absolute) convergence of the series for $e^{-x^2/2}$ in order to find this series. This is justified in Calculus 2 (MATH 1920) by the "Term-by-Term Integration Theorem"; see my online notes on Section 10.7. Power Series. We can then find

$$P(0 \leq Z \leq z) = \int_{0}^{z} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt$$

$$= \frac{1}{\sqrt{2\pi}} \left(\sum_{k=0}^{\infty} \frac{(-1)^k x^{2k+1}}{2^k(2k+1)k!} \right) \Bigg|_0^z = \frac{1}{\sqrt{2\pi}} \sum_{k=0}^{\infty} \frac{(-1)^k z^{2k+1}}{2^k(2k+1)k!}.$$

Notice that this is an alternating series and so can be estimated to any desired level of accuracy using the Alternating Series Estimation Theorem (see Theorem 15 in my online notes for Calculus 2 on Section 10.6. Alternating Series, Absolute and Conditional Convergence). We can use symmetry and the fact that $P(Z \le 0) = 1/2$ to find $P(Z \le z)$ for any $z \in \mathbb{R}$. Finally, we can compute $P(X \le x)$ where X has a $N(\mu, \sigma^2)$ distribution by computing "z-values" using the relationship $z = (x - \mu)/\sigma^2$.

Note. The cumulative distribution function of standard normal random variable Z is denoted $\Phi(z)$ and is

$$\Phi(z) = P(Z \le z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-t/2}\, dt.$$

The usual "z-table" used in introductory statistics is a table of $\Phi(z)$ values. Such a table is given in the text book in Table II of Appendix D. The table only gives values of $\Phi(z)$ for $z \ge 0$ since for $-z \le 0$ we have the identity $\Phi(-z) = 1 - \Phi(z)$ (which is to be shown in Exercise 3.4.1).

Example 3.4.2. To illustrate the use of Φ, let X have a $N(\mu, \sigma^2)$ distribution. Consider $P(\mu - \sigma < X < \mu + \sigma)$. We write this as $P(X < \mu + \sigma) - P(X \le \mu - \sigma)$. We need to convert to z-values and so consider $z = ((\mu + \sigma) - \mu)/\sigma = 1$ and $z = ((\mu - \sigma) - \mu)/\sigma = -1$, and so

$$P(\mu - \sigma < X < \mu + \sigma) = P(X < \mu + \sigma) - P(X \le \mu - \sigma) = \Phi(1) - \Phi(-1)$$

$$= \Phi(1) - (1 - \Phi(1)) = 2\Phi(1) - 1 \approx 2(0.8413) - 1 = 0.6826.$$

Similarly,

$$P(\mu - 2\sigma < X < \mu + 2\sigma) = 2\Phi(2) - 1 \approx 2(0.9772) - 1 = 0.9544$$

and

$$P(\mu - 3\sigma < X < \mu + 3\sigma) = 2\Phi(3) - 1 \approx 2(0.9987) - 1 = 0.9974.$$

This example illustrates "The 68-95-99.7 Rule" which one sees in introductory statistics; see my online notes for Introduction to Probability and Statistics (MATH 1530) on Chapter 3. The Normal Distributions. □

Example 3.4.4. All Moments of a Normal Distribution.

Recall from Section 1.9. Some Special Expectations, the mth moment of random variable X is $E(X^m)$. In Example 1.9.7 we calculated the mth moment of the standard normal distribution for all $m \in \mathbb{N}$. Let X be $N(\mu, \sigma^2)$. Then $X = \sigma Z + \mu$ where Z is $N(0,1)$. So for $k \in \mathbb{N}$ we have by the Binomial Theorem and the linearity of expectation (Theorem 1.8.2) we have

$$E(X^k) = E((\sigma Z + \mu)^k) = E\left(\sum_{j=0}^{k} \binom{k}{j} \sigma^j Z^j \mu^{k-j}\right) = \sum_{j=0}^{k} \binom{k}{j} \sigma^j E(Z^j) \mu^{k-j}.$$

We can now find the kth moment of X using the results of Example 1.9.7:

$$E(Z^j) = \begin{cases} 0 & \text{if } j \text{ is odd} \\ \frac{j!}{2^{j/2}(j/2)!} & \text{if } j \text{ is even.} \end{cases}$$

□

Note. We now find some common ground between normal distributions and χ^2 distributions.

Theorem 3.4.1. If the random variable X is $N(\mu, \sigma^2)$, where $\sigma^2 > 0$, then the random variable $V = (X - \mu)^2/\sigma^2$ is $\chi^2(1)$.

Note. The next theorem shows that a linear combination of independent normal random variables is itself a normally distributed.

Theorem 3.4.2. Let $X_1, X_2, \ldots, X_n$ be independent random variables such that, for $i = 1, 2, \ldots n$, X_i has a $N(\mu_i, \sigma_i^2)$ distribution. Let $Y = \sum_{i=1}^n a_i X_i$ where $a_1, a_2, \ldots, a_n$ are constants. Then the distribution of Y is $N\left(\sum_{i=1}^n a_i\mu_i, \sum_{i=1}^n a_i^2\sigma_1^2\right)$.

Note. If we take X_i, $i = 1, 2, \ldots, n$, to have the same distribution, then with each $a_1 = 1/n$ we can find the distribution of $\overline{X} = \sum_{i=1}^n X_i/n$ using Theorem 3.4.2, as follows.

Corollary 3.4.1. Let $X_1, X_2, \ldots, X_n$ be identical in distribution ("iid") random variables with a common $N(\mu, \sigma^2)$ distribution. Let $\overline{X} = n^{-1}\sum_{i=1}^n X_i$. Then $\overline{X}$ has a $N(\mu, \sigma^2/n)$ distribution.

Note. We now consider a sum of two normal distributions. We consider a standard normal random variable Z with high probability, but with a low probability "contamination" which we assume is normally distributed with mean 0 but with a larger variance $\sigma_c^2 > 1$. Specifically, define $I_{1-\varepsilon}$ as a discrete random variable satisfying:

$$I_{1-\varepsilon} = \begin{cases} 1 & \text{with probability } 1-\varepsilon \\ 0 & \text{with probability } \varepsilon \end{cases}$$

and assume Z and $I_{1-\varepsilon}$ are independent. Then let $W = ZI_{1-\varepsilon} + \sigma_c Z(1 - I_{1-\varepsilon})$. The cumulative distribution function of W is

$$\begin{aligned} F_W(w) &= P(W \le w) = P(W \le w \text{ and } I_{1-\varepsilon} = 1) + P(W \le w \text{ and } I_{1-\varepsilon} = 0) \\ &= P(W \le w \mid I_{1-\varepsilon} = 1)P(I_{1-\varepsilon} = 1) + P(W \le w \mid I_{1-\varepsilon} = 0)P(I_{1-\varepsilon}) \\ &= P(Z \le w)(1-\varepsilon) + P(Z \le w/\sigma_c)\varepsilon \text{ since } Z \text{ is } N(0,1) \\ &= \Phi(w)(1-\varepsilon) + \Phi(w/\sigma_c)\varepsilon. \end{aligned}$$

So W is a mixture of normal distributions. In Exercise 3.4.24, it is to be shown that $E(W) = 0$ and $\mathrm{Var}(W) = 1 + \varepsilon(\sigma_c^2 - 1)$. Differentiation F_W gives

$$F_W'(w) = f_W(w) = \Phi'(w)(1-\varepsilon) + \Phi'(w/\sigma_c)(\varepsilon/\sigma_c) = \varphi(w)(1-\varepsilon) + \varphi(w/\sigma_c)(\varepsilon/\sigma_c)$$

where φ is the probability density function of a standard normal (and $\Phi' = \varphi$ by Note 1.7.A).

Note. If we take $X = a + bW$ where $b > 0$ then we have $E(X) = a$ and $\mathrm{Var}(X) = b^2(1 + \varepsilon(\sigma_c^2 - 1))$ (based on the values of $E(W)$ and $\mathrm{Var}(W)$ given above). The cumulative distribution function of $X = a + bW$ is

$$F_X(x) = \Phi\left(\frac{x-a}{b}\right)(1-\varepsilon) + \Phi\left(\frac{x-a}{b\sigma_c}\right)\varepsilon.$$

This is a mixture of normal cumulative distribution functions. In Section 3.7 we further explore mixture distributions.

Revised: 7/21/2021