10 Normal Distribution

Described with two parameters μ and σ, the **normal distribution** is denoted as $X \sim N(\mu, \sigma^2)$. The probability density function of a normal distribution is given by

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

For a normal distribution, σ has a very precise meaning of spread of the distribution. There is a rule called "68-95-99.7%" which specifies how much probability there is between certain multiplicities of σ from the mean μ

$$\begin{aligned} P(\mu - \sigma < X < \mu + \sigma) &= 0.68 \\ P(\mu - 2\sigma < X < \mu + 2\sigma) &= 0.95 \\ P(\mu - 3\sigma < X < \mu + 3\sigma) &= 0.997 \end{aligned}$$

So the effective width of X is 6σ.

10.1 Standard Normal and Normal Table

Recall the probabilities for continuous random variables are obtained from areas under the p.d.f curve which require integration to get the c.d.f ($P(a < X \leq b) = P(b) - P(a)$). For the normal distribution there is no analytic formula for $F(x)$ and it requires numerical integration, that is areas obtained by tables or software.

Table 3 in the appendix and attached to this note is the special case for $\mu = 0$ and $\sigma = 1$, this special case is called the standard normal distribution denoted by $Z \sim N(0,1)$. The notation for the standard normal p.d.f is $\phi(z)$, c.d.f $\Phi(z)$. In table 3 we have values for $\Phi(z) \equiv P(Z \leq z)$ for $-3.9 \leq z \leq 3.9$.

To use table 3, find the first decimal of z on the left most column, then find the second decimal on the top row.

Example 4.11 Find $P(Z > 1.26)$, $P(Z > 0.86)$, and $P(-1.25 < Z < 0.37)$

Ans.

$$\begin{aligned} P(Z > 1.26) &= 1 - P(Z \leq 1.26) = 1 - 0.896165 = 0.103835 \\ P(Z > 0.86) &= 1 - \Phi(0.86) = 1 - 0.805106 = 0.194894 \\ P(-1.25 < Z < 0.37) &= \Phi(0.37) - \Phi(-1.25) = 0.644309 - 0.105650 = 0.538659 \end{aligned}$$

10.2 Standardizing a Normal Random Variable

Suppose that X is a normal random variable with $E(X) = \mu$ and $V(X) = \sigma^2$, that is $X \sim N(\mu, \sigma^2)$ then

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

Z is a normal random variable with $Z \sim N(0,1)$, that is Z is a standard normal random variable.

Z is called $Z - score$ or $Z - value$. It converts X into a dimensionless quantity (if it isn't already dimensionless) as the units in numerator and denominator cancel.

Suppose that $X \sim N(\mu, \sigma^2)$ and you want to find $F(x) = P(X \leq x)$. We write

$$P(X \leq x) = P\left(\frac{X - \mu}{\sigma} \leq \frac{x - \mu}{\sigma}\right) \tag{2}$$

$$= P\left(Z \leq \frac{x - \mu}{\sigma}\right) \tag{3}$$

So to find Normal probabilities or areas over a specified region, standardize the region to convert to a standard normal distribution and use table 3.

Example 4.13 Given $X \sim N(10, 4)$, find $P(X > 13)$

$$P(X > 13) = 1 - P\left(\frac{X - \mu}{\sigma} \leq \frac{13 - 10}{2}\right) = 1 - P(X \leq 1.5)$$
$$= 1 - 0.933193 = 0.06681$$

Example 4.15 Assume that in the detection of a digital signal, the background noise follows a normal distribution with a mean of 0 volts and a standard deviation of 0.45 volts. The system assumes a digital 1 has been transmitted with the voltage exceeds 0.9. What is the probability of detecting a digital signal 1 when none was sent?

Ans. Let the random variable N denote the noise voltage. The requested probability is

$$P(N > 0.9) = P\left(\frac{N}{0.45} > \frac{0.9}{0.45}\right) = P(Z > 2) = 1 - P(Z \leq 2) = 0.02275$$

This is the probability of a false positive. A false negative is when a signal but was not received (aka. What is the probability the signal was lower than the noise).

Suppose that $X \sim N(0, 0.45^2)$ is pure noise. Find a symmetric interval about 0 that would contain 99% of all possible observations. In other words, we want

$$P(-x \leq X \leq x) = P\left(\frac{-x - 0}{0.45} \leq \frac{X - 0}{0.45} \leq \frac{x - 0}{0.45}\right)$$
$$= P\left(\frac{-x}{0.45} \leq Z \leq \frac{x}{0.45}\right)$$
$$= 0.99$$

This examples utilize the standard normal tables in reverse. We are looking for an interval $(-z, z)$ under the Z curve which should equal 0.99. So the area outside $(-z, z)$ is 0.01 which should be split into two equal areas as by symmetry the 0.01 should be the area on either side of our distribution. So the tails of our distribution is 0.005. Therefore, if we look at the area from $(-\infty, z)$ we should have $0.005 + 0.99 = 0.995$. We now look for a z value such that $P(Z \leq z) = 0.995$, and we find $z = 2.58$. Now we solve for x

$$\frac{x}{0.45} = 2.58$$
$$x = 1.161$$

10.3 Percentiles

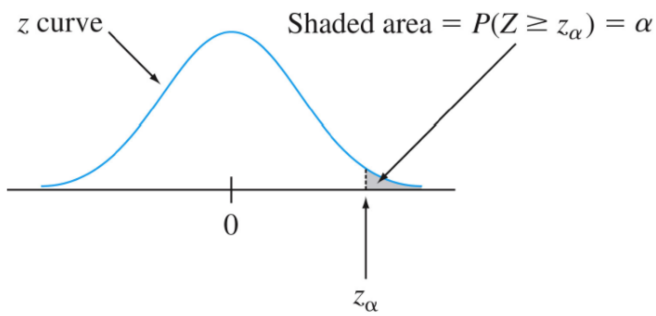
Let $0 < p < 1$. The 100^{th} percentile of X is a point x_p such that

$$P(X \leq x_p) \equiv F(x_p) = p \tag{4}$$

Which means the area to the left of x_p equals p

10.3.1 Percentile of Standard Normal Z

In statistics, for Normal Random Variables it is customary to refer to percentiles by area $\alpha \equiv 1 - p$ to the right of x_p. Thus for $0 < \alpha < 1$ let $z_\alpha = P(Z \geq \alpha)$ denote the $100(1-\alpha)^{th}$



From $Z = \dfrac{X - \mu}{\sigma}$ we can cross multiply to get

$$X = \mu + Z\sigma \tag{5}$$

$100p^{th}$ percentile of $X = \mu + \sigma(100p^{th})$ percentile of Z or more symbolically

$$x_p = \mu + \sigma z_{1-\alpha} \tag{6}$$

Example Coffee is dispensed by a machine is $X \sim N(8, 0.30^2)$ ounces. What container size c ensures that overflows occurs only 0.5% of the time?

Ans. Find c so that $P(X > c) = 0.005$. Since c requires to hold the 99.5^{th} percentile of coffee dispensed

$$c = 8 + 0.30(z_{1-\alpha})$$

$z_{1-\alpha}$ is given by $\alpha = 1 - 0.005 = 0.995$ (Recall, α is the area to the right). So we are looking for $z_{0.005} = 2.58$.

$$c = 8 + 0.30(2.58) = 0.77$$

A more easier way to pick $z_{1-\alpha}$ is to see that c must cover 99.5 percent. Thus we must find a $\Phi(z) = 0.995$