Lisa Yan and Jerry Cain   Lecture Notes #10
CS109   October 5, 2020

The Normal Distribution

Based on a chapter by Chris Piech

Pre-recorded lecture: Sections 1 and 2, definitions only.
In-lecture: Section 2's examples.
Not in lecture notes: Sampling demo. To be added soon.
Lecture 11: Section 3.

1   Normal Random Variable

The single most important random variable type is the Normal (aka Gaussian) random variable, parameterized by a mean (μ) and variance (σ^2). If X is a normal variable we write $X \sim \mathcal{N}(\mu, \sigma^2)$. The normal is important for many reasons: it is generated from the summation of independent random variables and as a result it occurs often in nature. Many things in the world are not distributed normally but data scientists and computer scientists model them as Normal distributions anyways. Why? Because it is the most entropic (conservative) distribution that we can apply to data with a measured mean and variance.

Properties

The Probability Density Function (PDF) for a Normal is:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

By definition a Normal has $E[X] = \mu$ and $\text{Var}(X) = \sigma^2$.

If X is a Normal such that $X \sim \mathcal{N}(\mu, \sigma^2)$ and Y is a linear transform of X such that $Y = aX + b$ then Y is also a Normal where $Y \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$.

There is no closed form for the integral of the Normal PDF, however since a linear transform of a Normal produces another Normal we can always map our distribution to the "Standard Normal" (mean 0 and variance 1) which has a precomputed Cumulative Distribution Function (CDF). The CDF of an arbitrary normal is:

$$F(x) = \Phi\left(\frac{x - \mu}{\sigma}\right)$$

Where Φ is a precomputed function that represents that CDF of the Standard Normal.

2 Projection to Standard Normal

For any Normal RV X we can find a linear transform from X to the **Standard Normal** $N(0, 1)$. That is, if you subtract the mean (μ) of the normal and divide by the standard deviation (σ), the result is distributed according to the standard normal (also called the unit Normal). We can prove this mathematically. Let $Z = \frac{X - \mu}{\sigma}$:

$$
\begin{aligned}
Z &= \frac{X - \mu}{\sigma} && \text{Transform } X \text{: subtract } \mu \text{ and divide by } \sigma \\
&= \frac{1}{\sigma} X - \frac{\mu}{\sigma} && \text{Use algebra to rewrite the equation} \\
&= aX + b && \text{define } a = \frac{1}{\sigma}, b = -\frac{\mu}{\sigma} \\
&\sim N(a\mu + b, a^2 \sigma^2) && \text{The linear transform of a normal is another normal} \\
&\sim N(\frac{\mu}{\sigma} - \frac{\mu}{\sigma}, \frac{\sigma^2}{\sigma^2}) && \text{Substitute values in for } a \text{ and } b \\
&\sim N(0, 1) && \text{The Standard Normal}
\end{aligned}
$$

An extremely common use of this transform is to express $F_X(x)$, the CDF of X, in terms of the CDF of Z, $F_Z(x)$. Since the CDF of the Standard Normal is so common, it gets its own Greek symbol, $\Phi(x)$.

$$
\begin{aligned}
F_X(x) &= P(X \leq x) \\
&= P\left(\frac{X - \mu}{\sigma} \leq \frac{x - \mu}{\sigma} \right) \\
&= P\left(Z \leq \frac{x - \mu}{\sigma} \right) \\
&= \Phi\left(\frac{x - \mu}{\sigma} \right)
\end{aligned}
$$

Why is this useful? Well, in the days when we couldn't call `scipy.stats.norm.cdf` (or on exams, when one doesn't have a calculator), people would look up values of the CDF in a table (see the last page of these notes). Using the Standard Normal means you only need to build a table of one distribution, rather than an indefinite number of tables for all the different values of μ and σ!

We also have an online calculator on the CS 109 website. You should learn how to use the Standard Normal table for the exams, however!

Example: Computing probabilities
Problem 1 : Let $X \sim N(3, 16)$, what is $P(X > 0)$?

$$
P(X > 0) = P\left(\frac{X - 3}{4} > \frac{0 - 3}{4} \right) = P\left(Z > -\frac{3}{4} \right) = 1 - P\left(Z \leq -\frac{3}{4} \right)
$$

$$
= 1 - \Phi\left(-\frac{3}{4} \right) = 1 - \left(1 - \Phi\left(\frac{3}{4} \right) \right) = \Phi\left(\frac{3}{4} \right) \approx 0.7734
$$

An alternative approach uses the idea that if F is the CDF of $X \sim \mathcal{N}(\mu, \sigma^2)$, then $F(x) = P\left(Z < \frac{x-\mu}{\sigma}\right) = \Phi\left(\frac{x-\mu}{\sigma}\right)$:

$$P(X > 0) = 1 - F(0) = 1 - \Phi(-3/4)$$
$$= 1 - (1 - \Phi(3/4)) = \Phi(3/4) \approx 0.7734.$$

Problem 2 : What is $P(2 < X < 5)$?

$$P(2 < X < 5) = P\left(\frac{2-3}{4} < \frac{X-3}{4} < \frac{5-3}{4}\right) = P\left(-\frac{1}{4} < Z < \frac{2}{4}\right)$$
$$= \Phi\left(\frac{2}{4}\right) - \Phi\left(-\frac{1}{4}\right) = \Phi\left(\frac{1}{2}\right) - \left(1 - \Phi\left(\frac{1}{4}\right)\right) \approx 0.2902$$

Alternative solution:

$$P(2 < X < 5) = F(5) - F(2) = \Phi\left(\frac{5-3}{4}\right) - \Phi\left(\frac{2-3}{4}\right)$$
$$= \Phi(1/2) - (1 - \Phi(1/4)) \approx 0.2902.$$

Problem 3 : What is $P(|X - 3| < 6)$?

$$P(|X - 3| > 6) = P(X < -3) + P(X > 9) = F(-3) + (1 - F(9)) = \Phi\left(\frac{-3-3}{4}\right) + \left(1 - \Phi\left(\frac{9-3}{4}\right)\right)$$
$$= \Phi(-3/2) + (1 - \Phi(3/2)) = 2(1 - \Phi(3/2)) \approx 0.1337.$$

Example: Wires and noise

Problem : You send voltage of 2 or -2 on a wire to denote 1 or 0. Let X = voltage sent and let R = voltage received. $R = X + Y$, where $Y \sim \mathcal{N}(0, 1)$ is noise. When decoding, if $R \geq 0.5$ we interpret the voltage as 1, else 0. What is $P(\text{error after decoding}|\text{original bit} = 1)$?

Solution : Given that we sent a 1, $X = 2$ and therefore $R = 2 + Y$. A decoding error occurs if we incorrectly interpret the signal as 0; this occurs if $R < 0.5$. Note that Y is the Standard Normal and therefore has CDF Φ:

$$P(R < 0.5|X = 2) = P(X + Y < 0.5|X = 2) = P(2 + Y < 0.5)$$
$$= P(Y < -1.5) = \Phi(-1.5) = 1 - \Phi(1.5) \approx 0.0668$$

Problem : What is $P(\text{error after decoding}|\text{original bit} = 0)$?

Solution : Given that we sent a 0, $X = -2$ and therefore $R = -2 + Y$. A decoding error occurs if we incorrectly interpret the signal as 1; this occurs if $R \geq 0.5$.

$$P(R \geq 0.5|X = -2) = P(X + Y \geq 0.5|X = -2) = P(-2 + Y \geq 0.5)$$
$$= P(Y \geq 2.5) = 1 - \Phi(2.5) \approx 0.0062.$$

Discussion : This example demonstrates an asymmetric decoding boundary, where there is lower probability of erroneously decoding a 0 as a 1 than vice versa. In many engineering circumstances, we may suffer stronger consequences if we turn something "on" when it was supposed to stay turned off. By setting the boundary of our decoding process asymmetrically, we can decrease the probability of this undesirable error.

3   Binomial Approximation

You can use a Normal distribution to approximate a Binomial $X \sim \text{Bin}(n, p)$. To do so define a normal $Y \sim \mathcal{N}(E[X], \text{Var}(X))$. Using the Binomial formulas for expectation and variance, $Y \sim \mathcal{N}(np, np(1 - p))$. This approximation holds for large n. Since a Normal is continuous and Binomial is discrete we have to use a continuity correction to discretize the Normal.

$$P(X = k) \approx P\left(k - \frac{1}{2} < Y < k + \frac{1}{2}\right) = \Phi\left(\frac{k - np + 0.5}{\sqrt{np(1 - p)}}\right) - \Phi\left(\frac{k - np - 0.5}{\sqrt{np(1 - p)}}\right)$$

Example: Website visitors

100 visitors to your website are given a new design. Let X = # of people who were given the new design and spend more time on your website. Your CEO will endorse the new design if $X \geq 65$. What is $P(\text{CEO endorses change}|\text{it has no effect})$?

$E[X] = np = 50$. $\text{Var}(X) = np(1 - p) = 25$. $\sigma = \sqrt{\text{Var}(X)} = 5$. We can thus use a Normal approximation: $Y \sim \mathcal{N}(50, 25)$.

$$P(X \geq 65) \approx P(Y > 64.5) = P\left(\frac{Y - 50}{5} > \frac{64.5 - 50}{5}\right) = 1 - \Phi(2.9) = 0.0019$$

Example: Stanford acceptance rate

Stanford accepts 2480 students and each student has a 68% chance of attending. Let X = # students who will attend. $X \sim \text{Bin}(2480, 0.68)$. What is $P(X > 1745)$?

$E[X] = np = 1686.4$. $\text{Var}(X) = np(1 - p) = 539.7$. $\sigma = \sqrt{\text{Var}(X)} = 23.23$. We can thus use a Normal approximation: $Y \sim \mathcal{N}(1686.4, 539.7)$.

$$P(X > 1745) \approx P(Y > 1745.5) = P\left(\frac{Y - 1686.4}{23.23} > \frac{1745.5 - 1686.4}{23.23}\right) = 1 - \Phi(2.54) = 0.0055$$