University of Regina

Statistics 252 – Mathematical Statistics

Lecture Notes

Winter 2005

Michael Kozdron

kozdron@math.uregina.ca
www.math.uregina.ca/∼kozdron

Contents

References

[1] Dennis D. Wackerly, William Mendenhall III, and Richard L. Schaeffer. *Mathematical Statistics with Applications*. Duxbury, Pacific Grove, CA, sixth edition, 2002.

Notation

The symbol $A := B$ means A *is defined to equal* B, whereas $C = D$ by itself means simply that C and D are equal. This is an important distinction because if you write $A := B$, then there is no need to verify the equality of A and B. They are equal by definition. However, if $C = D$, then there *IS* something that needs to be proved, namely the equality of C and D (which might not be obvious). Exercise 5.8 illustrates this subtle difference.

1 The Basic Idea of Statistics: Estimating Parameters

As you are no doubt aware from your previous statistics courses, the language of Statistics is probability. That is to say, although it is **TRIVIAL** and **BORING** to compute the summary statistics of a collection of numbers, it is **FASCINATING** to know that a random sample can be modelled as a collection of iid random variables from which quite **INTERESTING** ideas such as confidence intervals develop.

Recall that the overarching goal of Statistics is to estimate population *parameters*. This is done by calculating *statistics*, which are simply numbers computed from data, and using them as *point estimates* of the appropriate parameter.

As you learned in Stat 151, if you have a population with an unknown mean μ and unknown variance σ^2, and you randomly select a sample of data $\{y_1, y_2, \dots, y_n\}$, then an unbiased (point) estimator of μ is

$$\overline{y} := \frac{y_1 + \cdots + y_n}{n},$$

and a common unbiased (point) estimator of σ^2 is

$$s^2 := \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \overline{y})^2.$$

In Stat 252, we will study *many* of the properties of these (point) estimators.

Exercise 1.1. Being sure to carefully list all of the necessary assumptions, prove that

- $\overline{y}$ is an unbiased estimator of μ, and
- s^2 is an unbiased estimator of σ^2.

This is one reason for dividing by $n-1$ in the definition of the sample variance s^2, instead of dividing by (the more natural) n.

You will also recall that using only a single number (such as $\overline{y}$ or s^2) to estimate a parameter is not that beneficial because it is unlikely that point estimator will equal the parameter exactly. Instead, you learned that much more information is provided by a *confidence interval*.

Most confidence intervals that you have encountered are normal-distribution-based. This means that if $\hat{\theta}$ is a point estimator of a parameter θ, then

$$\hat{\theta} \pm z_\alpha \sqrt{\hat{V}(\hat{\theta})},$$

where $\hat{V}(\theta)$ is the (estimated) variance of $\hat{\theta}$, is a $(1-\alpha)\%$ confidence interval for θ.

You will also recall that the interpretation of a confidence interval is that *before* the sample is drawn,

$$P\left(|\theta - \hat{\theta}| \geq z_\alpha \sqrt{\hat{V}(\hat{\theta})}\right) = \alpha.$$

After the sample is drawn, no such probability statement is true. All you can do is simply *hope* that your confidence interval is not one of the unlucky $\alpha\%$.

2 Review of Random Variables

In the remainder of these notes, I outline what I believe are the basic and most fundamental concepts about random variables that every Stat 252 student ***MUST*** know. While the exposition in these notes varies from that in [1], the ideas are nonetheless the same. My suggestion to you is that you read these notes, and do the exercises *without* looking back at your Stat 251 material. Try and solve them using *only* what is given.

Suppose that Ω is the sample space of outcomes of an experiment.

Example 2.1. Flip a coin once: $\Omega = \{H, T\}$.

Example 2.2. Toss a die once: $\Omega = \{1, 2, 3, 4, 5, 6\}$.

Example 2.3. Toss a die twice: $\Omega = \{(i, j) : 1 \leq i \leq 6,\ 1 \leq j \leq 6\}$.

Note that in each case Ω is a *finite* set. (That is, the *cardinality* of Ω, written $|\Omega|$, is finite.)

Example 2.4. Consider a needle attached to a spinning wheel centred at the origin. When the wheel is spun, the angle ω made with the tip of the needle and the positive x-axis is measured. The possible values of ω are $\Omega = [0, 2\pi)$.

In this case, Ω is an *uncountably infinite* set. (That is, Ω is uncountable with $|\Omega| = \infty$.)

Definition 2.5. A *random variable* X is a function from the sample space Ω to the real numbers $\mathbb{R} = (-\infty, \infty)$. Symbolically, $X : \Omega \to \mathbb{R}$ via

$$\omega \in \Omega \mapsto X(\omega) \in \mathbb{R}.$$

Example 2.1 (continued). Let X denote the number of heads on a single flip of a coin. Then, $X(H) = 1$ and $X(T) = 0$.

Example 2.2 (continued). Let X denote the upmost face when a die is tossed. Then, $X(i) = i$, $i = 1, \ldots, 6$.

Example 2.3 (continued). Let X denote the sum of the upmost faces when two dice are tossed. Then, $X((i, j)) = i + j$, $i = 1, \ldots, 6$, $j = 1, \ldots, 6$. Note that the elements of Ω are ordered pairs, so that the function $X(\cdot)$ acts on (i, j) giving $X((i, j))$. We will often omit the inner parentheses and simply write $X(i, j)$.

Example 2.4 (continued). Let X denote the cosine of the angle made by the needle on the spinning wheel and the positive x-axis. Then $X(\omega) = \cos(\omega)$ so that $X(\omega) \in [-1, 1]$.

Remark. The use of the notation X and $X(\omega)$ is EXACTLY analogous to elementary calculus. There, the function f is described by its action on elements of its domain. For example, $f(x) = x^2$, $f(t) = t^2$, and $f(\omega) = \omega^2$ all describe EXACTLY the same function, namely, the function which takes a number and squares it.

Remark. For historical reasons, the term *random variable* (written X) is used in place of *function* (written f) and generic elements of the domain are denoted by ω instead of by x.

Remark. If X is a random variable, then we call $X(\omega)$ a *realization* of the random variable. The physical interpretation is that if X denotes the UNKNOWN outcome (*a priori*) of the experiment before it happens, then $X(\omega)$ represents the realization or observed outcome (*a posterior*) of the experiment after it happens.

Remark. It was A.N. Kolmogorov in the 1930's who formalized probability and realized the need to treat random variables as *measurable* functions. See Math 810: Analysis I or Stat 851: Probability.

3 Discrete and Continuous Random Variables

There are two extremely important classes of random variables, namely the so-called discrete and continuous. In a sense, these two classes are the same since the random variable is described in terms of a density function. However, there are slight differences in the handling of sums and integrals so these two classes are taught separately in undergraduate courses.

Important Observation. Recall from elementary calculus that the Riemann integral $\int_a^b f(x)\,dx$ is *defined* as an appropriate limit of Riemann sums $\sum_{i=1}^N f(x_i^*)\,\Delta x_i$. Thus, you are ALREADY FAMILIAR with the fact that SOME RELATIONSHIP exists between integrals and sums.

Definition 3.1. Suppose that X is a random variable. Suppose that there exists a function $f : \mathbb{R} \to \mathbb{R}$ with the properties that $f(x) \geq 0$ for all x, $\int_{-\infty}^{\infty} f(x)\,dx = 1$, and

$$P(\{\omega \in \Omega : X(\omega) \leq t\}) =: P(X \leq t) = \int_{-\infty}^{t} f(x)\,dx.$$

We call f the *(probability) density (function)* of X and say that X is a *continuous* random variable. Furthermore, the function F defined by $F(t) := P(X \leq t)$ is called the *(probability) distribution (function)* of X.

Fact. By the Fundamental Theorem of Calculus, $F'(x) = f(x)$.

Exercise 3.2. Prove the fact that $F'(x) = f(x)$, being sure to carefully state the necessary assumptions on f. Convince me that you understand the use of the dummy variables x and t in your argument.

Remark. There exist continuous random variables which *do not* have densities. For our purposes, though, we will always assume that our continuous random variables are ones with a density.

Example 3.3. A random variable X is said to be *normally distributed with parameters μ, σ^2*, if the density of X is

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\,\exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right), \quad -\infty < \mu < \infty,\ 0 < \sigma < \infty.$$

This is sometimes written $X \sim \mathcal{N}(\mu, \sigma^2)$. In Exercises 4.4 and 5.9, you will show that the mean of X is μ and the variance of X is σ^2, respectively.

Definition 3.4. Suppose that X is a random variable. Suppose that there exists a function $p : \mathbb{Z} \to \mathbb{R}$ with the properties that $p(k) \geq 0$ for all k, $\sum_{k=-\infty}^{\infty} p(k) = 1$, and

$$P(\{\omega \in \Omega : X(\omega) \leq N\}) =: P(X \leq N) = \sum_{k=-\infty}^{N} p(k).$$

We call p the *(probability mass function or) density* of X and say that X is a *discrete* random variable. Furthermore, the function F defined by $F(N) := P(X \leq N)$ is called the *(probability) distribution (function)* of X.

Example 2.3 (continued). If X is defined to be the sum of the upmost faces when two dice are tossed, then the density of X, written $p(k) := P(X = k)$, is given by

$p(2)$	$p(3)$	$p(4)$	$p(5)$	$p(6)$	$p(7)$	$p(8)$	$p(9)$	$p(10)$	$p(11)$	$p(12)$
1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

and $p(k) = 0$ for any other $k \in \mathbb{Z}$.

Remark. There do exist random variables which are neither discrete nor continuous; however, such random variables will not concern us.

Remark. If we know the distribution of a random variable, then we know all of the information about that random variable. For example, if we know that X is a normal random variable with parameters 0, 1, then we know everything possible about X without actually realizing it.

4 Law of the Unconscious Statistician

Suppose that $X : \Omega \to \mathbb{R}$ is a random variable (either discrete or continuous), and that $g : \mathbb{R} \to \mathbb{R}$ is a (piecewise) continuous function. Then $Y := g \circ X : \Omega \to \mathbb{R}$ defined by $Y(\omega) = g(X(\omega))$ is also a random variable.

We now define the expectation of the random variable Y, distinguishing the discrete and continuous cases.

Definition 4.1. If X is a discrete random variable and g is as above, then the *expectation* of $g \circ X$ is given by

$$\mathbb{E}(g \circ X) := \sum_k g(k)\, p(k)$$

where p is the probability mass function of X.

Definition 4.2. If X is a continuous random variable and g is as above, then the *expectation* of $g \circ X$ is given by

$$\mathbb{E}(g \circ X) := \int_{-\infty}^{\infty} g(x)\, f(x)\, dx$$

where f is the probability density function of X.

Notice that if $g(x) = 1$ for all x, then the expectation of X itself is

- $\mathbb{E}(X) := \sum_k k\, p(k)$, if X is discrete, and
- $\mathbb{E}(X) := \int_{-\infty}^{\infty} x\, f(x)\, dx$ if X is continuous.

Exercise 4.3. Suppose that X is a Bernoulli(p) random variable. That is, $P(X = 1) = p$ and $P(X = 0) = 1 - p$ for some $p \in [0, 1]$. Carefully verify that

- $\mathbb{E}(X) = p$,
- $\mathbb{E}(X^2) = p$, and
- $\mathbb{E}(e^{-\theta X}) = 1 - p(1 - e^{-\theta})$, for $0 \le \theta < \infty$.

Exercise 4.4. The purpose of this exercise is to make sure you can compute some straightforward (but messy) integrals. Suppose that $X \sim \mathcal{N}(\mu, \sigma^2)$; that is, X is a normally distributed random variable with parameters μ, σ^2. (See Example 3.3 for the density of X.) Show directly (without using any unstated properties of expectations or distributions) that

- $\mathbb{E}(X) = \mu$,
- $\mathbb{E}(X^2) = \sigma^2 + \mu^2$, and
- $\mathbb{E}(e^{-\theta X}) = \exp\left(-\theta\mu - \frac{\sigma^2\theta^2}{2}\right)$, for $0 \le \theta < \infty$.

Together with Exercise 5.9, this is the reason that if $X \sim \mathcal{N}(\mu, \sigma^2)$, we say that X is normally distributed with mean μ and variance σ^2.

5 Summarizing Random Variables

Definition 5.1. If X is a random variable, then the *mean* of X is the number $\mu := \mathbb{E}(X)$. Note that $-\infty \le \mu \le \infty$. If $-\infty < \mu < \infty$, then we say that X has a *finite mean*, or that X is an *integrable random variable*, and we write $X \in L^1$.

Example 5.2. Suppose that X is a Cauchy-distributed random variable. That is, X is a continuous random variable with density function

$$f(x) = \frac{1}{\pi} \cdot \frac{1}{x^2 + 1}.$$

Carefully show that $X \notin L^1$.

Definition 5.3. If X is a random variable with $\mathbb{E}(X^2) < \infty$, then we say that X has a *finite second moment* and write $X \in L^2$. If $X \in L^2$, then we define the *variance* of X to be the number $\sigma^2 := \mathbb{E}\left((X - \mu)^2\right)$. The *standard deviation* of X is the number $\sigma := \sqrt{\sigma^2}$. (As usual, this is the positive square root.)

Remark. It is an important fact that if $X \in L^2$, then it must be the case that $X \in L^1$. This follows from the so-called Cauchy-Schwarz Inequality. (See Exercises 5.18 and 5.19 below.)

Definition 5.4. If X and Y are both random variables in L^2, then the covariance of X and Y, written $\mathrm{Cov}(X,Y)$ is defined to be

$$\mathrm{Cov}(X,Y) := \mathbb{E}\left((X-\mu_X)(Y-\mu_Y)\right)$$

where $\mu_X := \mathbb{E}(X)$, $\mu_Y := \mathbb{E}(Y)$. Whenever the covariance of X and Y exists, we define the *correlation* of X and Y to be

$$\mathrm{Corr}(X,Y) := \frac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y} \tag{$\dagger$}$$

where σ_X is the standard deviation of X, and σ_Y is the standard deviation of Y.

Remark. By *fiat*, $0/0 := 0$ in ($\dagger$). Although this is sinful in calculus, we advanced mathematicians understand that such a decree is permitted as long as we recognize that it is only a *convenience* which allows us to simplify the formula. We need not bother with the extra conditions about dividing by zero. (See Exercise 5.20.)

Definition 5.5. We say that X and Y are *uncorrelated* if $\mathrm{Cov}(X,Y) = 0$ (or, equivalently, if $\mathrm{Corr}(X,Y) = 0$).

Theorem 5.6 (Linearity of Expectation). *Suppose that $X : \Omega \to \mathbb{R}$ and $Y : \Omega \to \mathbb{R}$ are (discrete or continuous) random variables with $X \in L^1$ and $Y \in L^1$. Suppose also that $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ are both (piecewise) continuous and such that $f \circ X \in L^1$ and $g \circ Y \in L^1$. Then, $f \circ X + g \circ Y \in L^1$ and, furthermore,*

$$\mathbb{E}(f \circ X + g \circ Y) = \mathbb{E}(f \circ X) + \mathbb{E}(g \circ Y).$$

Exercise 5.7. Prove the above theorem separately for both the discrete case and the continuous case. Be sure to state any assumptions or theorems from elementary calculus that you use.

Fact. If $X \in L^2$ and $Y \in L^2$, then the following computational formulæ hold:

- $\mathrm{Cov}(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$;
- $\mathrm{Var}(X) = \mathrm{Cov}(X,X) = \sigma^2$;
- $\mathrm{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$.

Exercise 5.8. Verify the three computational formulæ above.

Exercise 5.9. Using the third computational formula, and the results of Exercise 4.4, quickly show that if $X \sim \mathcal{N}(\mu, \sigma^2)$, then $\mathrm{Var}(X) = \sigma^2$. Together with Exercise 4.4, this is the reason that if $X \sim \mathcal{N}(\mu, \sigma^2)$, we say that X is normally distributed with mean μ and variance σ^2.

Definition 5.10. The random variables X and Y are said to be *independent* if $f(x,y)$, the joint density of (X,Y), can be expressed as

$$f(x,y) = f_X(x) \cdot f_Y(y)$$

where f_X is the density of X and f_Y is the density of Y.

Remark. Notice that we have combined the cases of a discrete and a continuous random variable into one definition. You can substitute the phrases *probability mass function* or *probability density function* as appropriate.

The following is an extremely deep, and important, result.

Theorem 5.11. *If X and Y are* independent *random variables with $X \in L^1$ and $Y \in L^1$, then*

- *the product XY is a random variable with $XY \in L^1$, and*
- $\mathbb{E}(XY) = \mathbb{E}(X)\,\mathbb{E}(Y)$.

Exercise 5.12. Using this theorem, quickly prove that if X and Y are independent random variables, then they are necessarily uncorrelated. (As the next exercise shows, the converse, however, is not true: there do exist uncorrelated, dependent random variables.)

Exercise 5.13. Consider the random variable X defined by $P(X = -1) = 1/4$, $P(X = 0) = 1/2$, $P(X = 1) = 1/4$. Let the random variable Y be defined as $Y := X^2$. Hence, $P(Y = 0|X = 0) = 1$, $P(Y = 1|X = -1) = 1$, $P(Y = 1|X = 1) = 1$.

- Show that the density of Y is $P(Y = 0) = 1/2$, $P(Y = 1) = 1/2$.
- Find the joint density of (X, Y), and show that X and Y are not independent.
- Find the density of XY, compute $\mathbb{E}(XY)$, and show that X and Y are uncorrelated.

Exercise 5.14. Prove Theorem 5.11 in the case when both X and Y are continuous random variables.

Exercise 5.15. Suppose that $X : \Omega \to \mathbb{R}$ and $Y : \Omega \to \mathbb{R}$ are independent, integrable, continuous random variables with densities f_X, f_Y, respectively. Let $g : \mathbb{R} \to \mathbb{R}$ and $h : \mathbb{R} \to \mathbb{R}$ be continuous functions such that $g \circ X \in L^1$ and $h \circ Y \in L^1$. Prove that $\mathbb{E}\left((g \circ X) \cdot (h \circ Y)\right) = \mathbb{E}(g \circ X)\,\mathbb{E}(h \circ Y)$.

As a consequence of the previous exercise, we have the following very important result.

Theorem 5.16 (Linearity of Variance when Independent). *Suppose that $X : \Omega \to \mathbb{R}$ and $Y : \Omega \to \mathbb{R}$ are (discrete or continuous) random variables with $X \in L^2$ and $Y \in L^2$. If X and Y are* independent*, then*

$$\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y).$$

It turns out that Theorem 5.11 is not quite true when X and Y are not independent. However, the following is a probabilistic form of the ubiquitous Cauchy-Schwarz inequality, and usually turns out to be good enough.

Theorem 5.17 (Cauchy-Schwarz Inequality). *Suppose that X and Y are both random variables with finite second moments. That is, $X \in L^2$, and $Y \in L^2$. It then follows that*

- *the product XY is a random variable with $XY \in L^1$, and*
- $(\mathbb{E}(XY))^2 \le \mathbb{E}(X^2)\,\mathbb{E}(Y^2)$, *and*
- $(\mathrm{Cov}(X, Y))^2 \le \mathrm{Var}(X)\ \mathrm{Var}(Y)$.

Exercise 5.18. Using the first part of the Cauchy-Schwarz inequality, show that if $X \in L^2$, then $X \in L^1$.

Exercise 5.19. Using the second part of the Cauchy-Schwarz inequality, show that if $X \in L^2$, then $X \in L^1$.

Exercise 5.20. Using the third part of the Cauchy-Schwarz inequality, you can now make sense of the Remark following Definition 5.4. Show that if X and Y are random variables with $\operatorname{Var}(X) = \operatorname{Var}(Y) = 0$, then $\operatorname{Cov}(X, Y) = 0$.

The following facts are also worth mentioning.

Theorem 5.21. *If $a \in \mathbb{R}$ and $X \in L^2$, then $aX \in L^2$ and $\operatorname{Var}(aX) = a^2 \operatorname{Var}(X)$. In particular, $\operatorname{Var}(-X) = \operatorname{Var}(X)$.*

Theorem 5.22. *If $X_1, X_2, \ldots, X_n$ are L^2 random variables, then*

$$\operatorname{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} \operatorname{Cov}(X_i, X_j).$$

In particular, if $X_1, X_2, \ldots, X_n$ are uncorrelated *L^2 random variables, then*

$$\operatorname{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \operatorname{Var}(X_i).$$

6 The Big Theorems

The *Grand Prix* of probability theory, and perhaps its most important theorem, is the Central Limit Theorem. The importance of this remarkable result cannot be understated. In fact, it should be called the *Fundamental Theorem of Probability.*

Question. Did you ever wonder what was so *fundamental* about the Fundamental Theorem of Calculus?

Question. Did you ever wonder what was so *normal* about a normal distribution?

As a warm-up, here is a form of the poorly-named *Law of Averages*. Suppose that $\{X_n\}$, $n = 0, 1, 2, \ldots$, is a collection of independent, and identically distributed random variables. Suppose further that there exist constants $\mu \in (-\infty, \infty)$ and $\sigma \in (0, \infty)$ such that for each n, the random variable $X_n \in L^2$ with $\mathbb{E}(X_n) = \mu$ and $\operatorname{Var}(X_n) = \sigma^2$.

For each n, let M_n be the random variable defined by

$$M_n := \frac{X_1 + \cdots + X_n}{n}.$$

Then, obviously, for each n, we have $\mathbb{E}(M_n) = \mu$ and $\operatorname{Var}(M_n) = \sigma^2/n$. However, we can also say something about the *limiting values* of M_n, $\mathbb{E}(M_n)$, and $\operatorname{Var}(M_n)$ as $n \to \infty$.

Exercise 6.1. Prove that for each n,

- $\mathbb{E}(M_n) = \mu$, and
- $\mathrm{Var}(M_n) = \dfrac{\sigma^2}{n}$.

Think back to Exercise 1.1. Hmmm, they are the same problem!

Theorem 6.2 (Strong Law of Large Numbers). *With M_n defined as above, the following limits hold:*

- $\lim_{n\to\infty} \mathbb{E}(M_n) = \mu$, *and*
- $\lim_{n\to\infty} \mathrm{Var}(M_n) = 0$, *and*
- $P\left\{\omega : \lim_{n\to\infty} M_n(\omega) = \mu\right\} = 1$.

Remark. The third conclusion of the Strong Law of Large Numbers makes *intuitive* sense. If X is a random variable with mean μ and variance 0, then it must be constant, namely,

$$P\{\omega : X(\omega) = \mu\} = 1.$$

This follows from Chebychev's (Tchebysheff's) inequality.

Similarly, if M_n has constant mean, and limiting variance 0, then we are not surprised that the limiting random variable is constant. *However*, this fact is not quite so obvious, which is why a proof is required.

And here it is ...

Theorem 6.3 (Central Limit Theorem). *Suppose that $\{X_n\}$, $n = 0, 1, 2, \ldots$, is a collection of independent, and identically distributed random variables. Suppose further that there exist constants $\mu \in (-\infty, \infty)$ and $\sigma \in (0, \infty)$ such that for each n, the random variable $X_n \in L^2$ with $\mathbb{E}(X_n) = \mu$ and $\mathrm{Var}(X_n) = \sigma^2$. For each n, let S_n be the random variable defined by*

$$S_n := \frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}}.$$

Then, for $x \in \mathbb{R}$, it follows that as $n \to \infty$,

$$P(S_n \le x) \to \Phi(x) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}}\, dt.$$

That is, the limiting distribution of S_n as $n \to \infty$ is a standard normal.

Remark. The hypotheses about a collection of iid random variables say NOTHING about their distribution. They could share ANY distribution (be it discrete or continous, or a mixture of the two) so long as they have a finite second moment.

Remark. While the SLLN says something about the limiting value of the means and variances, the CLT says something about the limiting *distribution*. Knowing that a random variable has mean 0 and variance 1 doesn't say much. Knowing that a random variable is $\mathcal{N}(0, 1)$ says *everything*!