FUNCTIONS OF RANDOM VARIABLES

Video 1. Start with this example connecting statistics to probability

Theorem 1. *Let $g(x)$ be a continuous function and let X be a random variable.*

- *If X is a continuous RV with PDF $f(x)$, then $E[g(X)] = \int_{-\infty}^{-\infty} g(x)f(x)\ dx$*
- *If X is a discrete RV with PMF $f(x)$, then $E[g(X)] = \sum_x g(x)f(x)$*

1. Let X be a continuous random variable with mean 50 and standard deviation 4. Calculate $E\left(\frac{X-50}{4}\right)$ and $\mathrm{Var}\left(\frac{X-50}{4}\right)$

Video 2. Once you're done or stuck, watch the lesson (in which the following theorem is proved).

Theorem 2. *Let X be a random variable and let a and b be constants. Then $E(aX+b) = aE(X)+b$ and $Var(aX+b) = a^2\, Var(X)$.*

Theorem 3. *Let $X_1, X_2, \ldots, X_n$ be any random variables. Then $E(X_1 + X_1 + \cdots + X_n) = E(X_1) + E(X_2) + \cdots + E(X_n)$.*

Theorem 4. *Let $X_1, X_2, \ldots, X_n$ be independent random variables. Then $Var(X_1 + X_2 + \cdots + X_n) = Var(X_1) + Var(X_2) + \cdots + Var(X_n)$.*

Definition. We say that the random variables $X_1, X_2, \ldots, X_n$ are a **random sample** if they are independent and identically distributed. The **sample size** is n. Their common distribution is called the **population distribution**. Some **sample statistics**:

(1) The **sample total**: $T = \sum_{i=1}^{n} X_i$

(2) The **sample mean**: $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

(3) The **sample variance**: $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$

Our goal is to understand the distributions of these sample statistics.

2. Let $X_1, X_2, \ldots, X_n$ be a random sample from a population with mean μ and variance σ^2. Calculate the expected value and variance of the sample mean (that is, $E(\overline{X})$ and $\mathrm{Var}(\overline{X})$).

Date: March 9, 2020.

Video 3. Watch the next video.

Theorem 5 (Central Limit Theorem). *If $X_1, X_2, \ldots, X_n$ comprise a random sample from a population with mean μ and variance σ^2, then the limiting distribution (as $n \to \infty$) of $\dfrac{\overline{X} - \mu}{\sigma/\sqrt{n}}$ is a standard normal distribution.*

3. Continuing with the spoke example from the video, suppose another kind of spoke has a mean mass of 7.5g with a standard deviation of 0.6g. These spokes are stronger, so you only need to use 32 per wheel.

a) On average, which spokes make a lighter wheel (these or the spokes in the video)?

b) What is the probability that the total mass of the spokes in a wheel exceeds 245g if you use these spokes?

4. Suppose you roll a die 64 times. Use the CLT to estimate the probability that the sum of all the rolls is at least 200. Hint: think of this as a random sample of size 64 from an underlying population with the mean and variance of a single roll of the die (you may need to calculate this mean and variance).

5. If the underlying population is normally distributed, then $\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$ is a standard normal random variable (not just approximately standard normal), so our calculations are the same, but our answers are exact and **all sample sizes work**. Suppose we have a normally distributed population with mean 100 and standard deviation 10. Let $\overline{X}$ be the mean of a random sample of size n. Calculate $P(98 < \overline{X} < 102)$ for each sample size.

a) $n = 1$

b) $n = 25$

c) $n = 100$

d) $n = 10000$

6. As in the coronavirus example, binomial distributions show up when we count the number of successes in a random sample. For example, suppose you roll a die 64 times. Let T be the number of sixes you roll.

a) What is the (exact) distribution of T?

b) Calculate $P(T \geq 12)$.

c) The CLT says that T is approximately normal with mean $64(1/6)$ and variance $64(1/6)(5/6)$. Use this normal approximation to estimate $P(T \geq 12)$. How does this answer compare with the exact answer (in part b)?