I. Law of large numbers & central limit theorem

1. Let $X_1, X_2, \ldots, X_n$ be Binomial IIDRVs counting the number of heads after flipping a fair coin twice. Let $\overline{X} = \frac{X_1+\cdots+X_n}{n}$ be the average of the first n X_i's, and Z the normalized RV corresponding to $\overline{X}$. For $n = 5$:

(a) List the range $R_{\overline{X}}, \overline{\mu}$, and $\overline{\sigma}$. Compute the p.m.f. $f_{\overline{X}}$ and (optional) sketch its graph.

> As always, we have the mean $\overline{\mu}$ of $\overline{X}$ as the expected value of any individual X_i, so
>
> $$\overline{\mu} = E(X_i) = 2 \cdot \frac{1}{2} = 1.$$
>
> We also have a formula for $\overline{\sigma}$:
>
> $$\overline{\sigma} = \frac{\sigma}{\sqrt{n}} = \frac{\sqrt{2 \cdot \frac{1}{2} \cdot \frac{1}{2}}}{\sqrt{5}} = \frac{1}{\sqrt{10}}.$$
>
> For the range, we note that $X_1 + X_2 + X_3 + X_4 + X_5$ takes values from 0 to 10, and looks like a Binomial DRV with $n = 10, p = 1/2$. So $R_{\overline{X}} = \{\frac{0}{5}, \frac{1}{5}, \ldots, \frac{10}{5}\}$, and the PMF is as follows:

$\frac{0}{5}$	$\frac{1}{5}$	$\frac{2}{5}$	$\frac{3}{5}$	$\frac{4}{5}$	$\frac{5}{5}$	$\frac{6}{5}$	$\frac{7}{5}$	$\frac{8}{5}$	$\frac{9}{5}$	$\frac{10}{5}$
$\frac{\binom{10}{0}}{2^{10}}$	$\frac{\binom{10}{1}}{2^{10}}$	$\frac{\binom{10}{2}}{2^{10}}$	$\frac{\binom{10}{3}}{2^{10}}$	$\frac{\binom{10}{4}}{2^{10}}$	$\frac{\binom{10}{5}}{2^{10}}$	$\frac{\binom{10}{6}}{2^{10}}$	$\frac{\binom{10}{7}}{2^{10}}$	$\frac{\binom{10}{8}}{2^{10}}$	$\frac{\binom{10}{9}}{2^{10}}$	$\frac{\binom{10}{10}}{2^{10}}$

(b) Which values of $\overline{X}$ are solutions to the inequality $|\overline{X} - \overline{\mu}| > \frac{2}{3}$?

> We must either have $\overline{X} > \overline{\mu} + \frac{2}{3} = \frac{5}{3}$ or $\overline{X} < \overline{\mu} - \frac{2}{3} = \frac{1}{3}$. So $\overline{X}$ must take values in the set $\{\frac{0}{5}, \frac{1}{5}, \frac{9}{5}, \frac{10}{5}\}$.

(c) Compute the probability $P(|\overline{X} - \overline{\mu}| > \frac{2}{3})$.

> We simply sum the corresponding probabilities to get $\frac{1+10+10+1}{2^{10}} = \frac{22}{2^{10}}$.

(d) What is the probability $P(|\overline{X} - \overline{\mu}| \leq \frac{2}{3})$?

> This is the complement of the previous probability, so is $\frac{2^{10}-22}{2^{10}} \approx 0.9785$.

(e) Rewrite the inequality $|\overline{X} - \overline{\mu}| \leq \frac{2}{3}$ into inequalities of the form $a \leq Z \leq b$ for the normalized RV Z.

> If $\overline{X} - \overline{\mu} \leq \frac{2}{3}$ then dividing by $\overline{\sigma}$ gives us $Z \leq \frac{2}{3\overline{\sigma}} = \frac{2\sqrt{10}}{3}$. If $\overline{\mu} - \overline{X} \leq \frac{2}{3}$ then dividing by $\overline{\sigma}$ and negating gives us $Z \geq -\frac{2}{3\overline{\sigma}} = -\frac{2\sqrt{10}}{3}$.

(f) Using the standard normal table, approximate $P(a \leq Z \leq b)$ for the two numbers a and b you found above. What is the z-score you used in the table?

We approximate $\frac{4\sqrt{5}}{3} \approx 2.11$, which has corresponding value in the table 0.4826. Doubling gives us the estimate 0.9652.

(g) What is the relation between $P(|\overline{X} - \overline{\mu}| \leq \frac{2}{3}), P(a \leq Z \leq b)$, and the number you got in part (f)? Are they equal? Approximately equal? Why?

The first two probabilities are exactly the same, as the inequalities can be obtained from each other. The number we got in part (f) is an approximation to the first two probabilities, which we get by replacing the normalized variable Z with the standard normal curve. The Central Limit Theorem says that this is a reasonable thing to do for n large.

(h) Imagine we have done the same thing for $n = 10$. How would the probability we computed in part (d) change?

The Law of Large Numbers says that the probability in part (d) should go to 1 as n increases to infinity. So that probability should be larger if $n = 10$. If you actually compute it, it is

$$1 - \frac{1 + 20 + \binom{20}{2} + \binom{20}{3} + \binom{20}{4}}{2^{19}} \approx 99.74\%.$$

II. PDFs and continuous random variables

2. For each of the following, check whether the given function $f(x)$ is a valid PDF, and compute $P(0 < X < 1)$.

(a) $f(x) = 2x^2$ for x between 0 and 1, and 0 otherwise.

(b) $f(x) = xe^{x^2}$ for x between 0 and 1, and 0 otherwise.

(c) $f(x) = \cos x$ for x between 0 and π, and 0 otherwise.

(d) $f(x) = \cos x$ for x between 0 and $\frac{\pi}{2}$, and 0 otherwise.

(e) $f(x) = \frac{1}{x^2}$ for $x \geq 1$, and 0 otherwise.

(a) We note that $f(x)$ is certainly a non-negative function from the real numbers to the real numbers, so we need to check that its overall integral is 1. Well,

$$\int_0^1 3x^2 dx = \frac{3x^3}{3}|_0^1 = 1 - 0 = 1.$$

So this is a valid PDF. The probability $P(0 < X < 1)$ is $\int_0^1 f(x)dx = 1$.

(b) Again, this is certainly a non-negative function from the real numbers to the real numbers. We compute the integral

$$\int_0^1 xe^{x^2} dx = \frac{1}{2}\int_0^1 e^u du = \frac{1}{2}\left(e^u|_0^1\right) = \frac{e-1}{2}.$$

This is therefore not a valid PDF.

(c) We recall that $\cos x$ is negative for $\pi/2 < x < \pi$, so $f(x)$ is not nonnegative. Hence this is not a valid PDF.

(d) This is a nonnegative function, so we need to check its integral. We have

$$\int_0^{\pi/2} \cos x dx = \sin x|_0^{\pi/2} = 1 - 0 = 1.$$

So this is a valid PDF. We have the probability $P(0 < X < 1)$ as

$$\int_0^1 \cos x dx = \sin x|_0^1 = \sin 1 - 0 = \sin 1.$$

(e) Again, this is a nonnegative function, so we compute:

$$\int_0^1 f(x)dx = \int_1^\infty \frac{1}{x^2}dx = -\frac{1}{x}|_1^\infty = -(0-1) = 1.$$

So this is a valid PDF. The probability $P(0 < X < 1)$ is just

$$\int_0^1 f(x)dx = \int_1^1 f(x)dx = 0.$$

3. For each of the functions in the previous problem that were not PDFs, how might you turn them into valid PDFs?

We can rescale the function in part (b) by $\frac{2}{e-1}$, which will fix the integral. For the function in part (c), we could (for example) either restrict the range as in part (d) or take the absolute value of the function and then rescale.

4. Let $f(t)$ denote the PDF for the time in minutes it takes to get from Berkeley to San Francisco by BART. Write the probability that your trip takes less than forty minutes as an integral.

This is $\int_{-\infty}^{40} f(x)dx$. Assuming $f(x) = 0$ if $x \leq 0$, we could also write this as $\int_0^{40} f(x)dx$.