Unit 1 Introduction

Lecture 1

Probability review: dice rolling game

1/1 point (graded)

Alice and Bob play a game where two fair six-sided dice are rolled.

Alice gets $1 if the sum of the numbers of the two dice is a prime number. (The number 1 is not prime.)

Bob gets $3 if the numbers on the two dice are the same (e.g. 1-1, 2-2, ...).

Who makes more money on average (i.e. in expectation)?

- ◯ Alice
- ◉ Bob
- ◯ It does not matter.
- ◯ Not enough information to decide.

✔

Solution:

The set of possible outcomes is $\{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$, of which the prime numbers are $\{2, 3, 5, 7, 11\}$. Bob wins $3 whenever he sees one of six outcomes, out of a total of 36. This means he will earn $\frac{6}{36} \times 3 = \frac{1}{2}$ dollars per game in expectation. On the other hand, Alice wins $1 when she sees one of fifteen outcomes. Be careful here, because there are multiple ways for two six-sided dice to sum to 3,5,7 or 11 (As opposed to 2, for which there is only one possible roll). Therefore, Alice wins $1 \times \frac{15}{36}$ dollars in expectation.

Overall, Bob makes more money on average.

Probability or statistics 1

2/2 points (graded)

Determine whether each of the problems below is a probabilistic or a statistical problem. (You are **not** asked to solve them.)

1. Assume we have a population consisting of two subpopulations, A and B. A particular drug has a different chance of treatment success depending on the subpopulation, namely 70% for group A and 50% for group B.

   Assume that subgroup A is 10% of the entire population and subgroup B is 90%. What is the chance of a successful treatment if we pick a random person from the entire population?

   This is a

   - ◯ statistical problem
   - ◉ probabilistic problem

   ✔

2. Now, consider the scenario where we do not know the true composition of the population, which may be different from the previous setup. Among 1000 randomly chosen patients, we observe that the treatment was successful in 700 of them. What is a good estimate of the composition of the population?

   This is a

   - ◉ statistical problem
   - ◯ probabilistic problem

   ✔

Solution:

1. The first one is a probabilistic problem, because we are given all relevant parameters and are trying to compute corresponding derived probabilities. In particular, if we denote the subpopulation of a randomly selected person by $X \in \{A, B\}$ and the treatment outcome of drug D by $Y \in \{\text{sucess, failure}\}$, we are given $\mathbf{P}(X = A)$, $\mathbf{P}(X = B)$, $\mathbf{P}(Y = \text{success}|X = A)$, and $\mathbf{P}(Y = \text{success}|X = B)$ and are asked to compute $\mathbf{P}(Y = \text{success})$.

2. The second one is a statistical problem, because we are trying to estimate an underlying probabilistic parameter from data. More explicitly, we have 1000 i.i.d. draws from the Bernoulli random variable Y, 700 of which correspond to $Y = \text{success}$, and are now being asked to draw conclusions about $\mathbf{P}(Y = \text{success})$ and from there about $\mathbf{P}(X = A)$ and $\mathbf{P}(X = B)$.

Probability or statistics 2

2/2 points (graded)

John Arbuthnot wrote a paper in 1710 entitled 'An Argument for Divine Providence', where he studied, based on the Christening records in London for 1629-1710, the chances that a randomly chosen baby born is a girl or a boy. Is this a statistical problem, or a probabilistic problem?

- (x) A statistical problem.
- ( ) A probabilistic problem.

✔

Next, you read Arbuthnot's paper, and went to a gyneacology facility, in which there are 10 babies whom are expected to be born on the day you arrived, and you are interested in, what are the odds that 6 of those will be a boy, and the remaining will be a girl. Is this a statistical problem, or a probabilistic problem?

- ( ) A statistical problem.
- (x) A probabilistic problem.

✔

Solution:

The first one is a statistical problem, and the second one is a probabilistic problem. To see this, suppose that each newborn baby is a boy with probability p, and a girl with probability $1 - p$. Suppose also that the sex of each newborn is independent of the sex of all others. The data that Arbuthnot analyzed simply corresponds to realizations of this Bernoulli variable; and from that knowledge, we simply want to extract the underlying parameter, p. This is an example of a statistical problem.
You take Arbuthnot's finding, and assume that this is the 'true' probability for the aforementioned birth process; and want to compute a probability, which is simply

$$\binom{10}{6} p_A^6 (1 - p_A)^4,$$

where p_A is the value that Arbuthnot has reported (namely, you are computing the probability that a certain Binomial random variable is equal to 6.

Probability or statistics 3

2/2 points (graded)

A doctor realizes that there is an allergy medicine which is effective in treating seasonal allergies with probability at least 90%. From here, he claims:

- Out of 100 patients admitted to clinic with seasonal allergies, this drug will cure 90 patients, on average.
- At least 70 patients will be cured, with 99.99% chance.

Does he rely on statistics, or probability?

- ○ Statistics
- ◉ Probability

✔

Now, a newly-hired scientist at a pharmacology company performs an experiment, and based his observations, deduces that, "I am 95% confident that if we repeat this experiment, then the drug will be effective on between 85% and 95% patients." Does he rely on statistics, or probability?

- ◉ Statistics
- ○ Probability

✔

Solution:

- The doctor relies on probability. The point of discussion is about averages and the odds that at least 70 patients will be cured.
- The scientist relies on statistics, since is using observations This is hinted at, because he discusses confidence regions.

Review: probability question

0/1 point (graded)

Assume that we observe three draws, X_1, X_2, X_3 from a Bernoulli distribution with parameter $p = \frac{1}{2}$. For example, imagine that in the model for the preferred head direction for kissing, either direction were actually equally likely and we observed three kissing couples.

What is the probability of observing at least two ones, i.e., what is $\mathbf{P}\left(\sum_{i=1}^{3} X_i \geq 2\right)$?

(1/2)^3 ✖ **Answer:** 0.5

Solution:

$\sum_{i=1}^{3} X_i$ follows a Binomial distribution with parameters $n = 3$ and $p = \frac{1}{2}$, hence the probability in question is

$$\mathbf{P}\left(\sum_{i=1}^{3} X_i \geq 2\right) = \binom{3}{2}\left(\frac{1}{2}\right)^3 + \binom{3}{3}\left(\frac{1}{2}\right)^3 = \frac{4}{8} = \frac{1}{2}.$$

Confidence, continued

1/1 point (graded)

If in the model above, let us assume we decided to consider two or more right-turns as significant evidence for a predisposition of this direction for kissing. Now, 10 students go out and each observe three different couples kissing. How many of them would on average come to the conclusion that right-leaning is more common than left-leaning when kissing?

5 ✔ **Answer:** 5

Solution:

We just computed the chance for one of these events to occur to be $\frac{1}{2}$, so if we perform 10 repeats, we expect it to happen 5 times.

Friendships

1/1 point (graded)

In a group of n people indexed 1 through n, each pair (i, j) (there are $\binom{n}{2}$ of them) are either friends, or not friends. To model this situation, we assign a random variable to each pair. Which one of the probability distributions below is the most appropriate model?

- ( ) A Poisson random variable.
- ( ) A Gaussian random variable.
- (•) A Bernoulli random variable.
- ( ) An exponential random variable.

✔

Solution:

We need a random variable, which takes only two values (for convenience, 1 for being friends; and 0 for strangers), and this is precisely a Bernoulli distribution.

Friendships (Continued)

1/2 points (graded)

With the setup as in the problem above, we say a group of four people is "interesting", if there are at most five pairs who are friends. Assume that each pair of people are friends, independent of every other pair, with probability $1/2$. Let N be the number of pairs that are friends in this group.

- What distribution does N follow?

  - ( ) Poisson
  - ( ) Bernoulli
  - (•) Binomial

  ✔

- What is the probability that a randomly chosen group of four people is "interesting"? (Enter your answer as a fraction (recommended) or enter as a decimal accurate to nearest 0.001.)

$\mathbb{P}(N \leq 5) =$ 1/2 ✖ **Answer:** 63/64

Solution:

- There are, in total, $\binom{4}{2} = 6$ different pairs. Notice that N is a random sum of 6 Bernoulli trials; this is a binomial random variable: $N \sim \text{Bin}(6, 1/2)$.
- $\mathbb{P}(N \leq 5) = 1 - \mathbb{P}(N = 6) = 1 - (1/2)^6$.

How many interesting groups?

1/1 point (graded)

Following the model above, if 128 different people each observe one randomly chosen groups of four people, how many times on average do these observations lead to the conclusion that the person's chosen group is interesting?

126 ✔ **Answer:** 126

Solution:

We just computed the chance for one of these events to occur to be $1 - 1/64$, so if we perform 128 repeated experiments, we expect it to happen $128\,(1 - \frac{1}{64}) = 126$ times.

Independence

1/1 point (graded)

Consider a probabilistic experiment where we roll a dice and toss a coin. We compute the probability that the fair dice gives 5 and the fair coin lands Heads. What assumptions are we implicitly using in this specific calculation: $\Pr(5,\text{Heads}) = \Pr(5) \cdot \Pr(\text{Heads}) = \frac{1}{6} \cdot \frac{1}{2}$? Choose all that apply, so that the chosen assumptions best capture the required concepts.

- [x] Each dice roll is uniformly distributed within the set $\{1, 2, 3, 4, 5, 6\}$ and each coin toss is uniformly distributed in $\{\text{Heads}, \text{Tails}\}$.
- [x] The dice roll and coin toss are independent.
- [ ] The random variables corresponding to outputs of each of these experiments are i.i.d.

✔

Solution:

The correct answers are the first and second choices.

Let X denote the output of the dice roll and Y denote the output of the coin toss. We are looking at the probability

$$\begin{aligned}\mathbb{P}(X = 5, Y = \{Heads\}) &= \mathbb{P}(X = 5)\,\mathbb{P}(Y = \{Heads\}) \\ &= \frac{1}{6} \cdot \frac{1}{2}.\end{aligned}$$

The first line, where we express the joint probability as a product, uses the fact that coin toss and dice roll are independent. The second line, where we substitute the values $1/6$ and $1/2$, uses the uniformity assumption to explicitly compute these probabilities.

Remember from the course *Probability–the Science of Uncertainty and Data* that **i.i.d.** stands for **independent and identically distributed** .

A collection of random variables $X_1, \ldots, X_n$ are **i.i.d.** if

1. each X_i follows a distribution $\mathbf{P}_i$, all those distributions $\mathbf{P}_i$ are the same, and
2. X_i are (mutually) independent

Decide which of the following collections are (approximately) i.i.d. (independent and identically distributed).
(Choose all that apply.)

- [ ] People selected randomly (with replacement) by their address from a directory. ✔
- [x] The first two consecutive words of a random page in a book.
- [x] Repeated dice rolls of the same die. ✔
- [x] Temperature measurements on Monday and Tuesday in the same week.

✖

Solution:

If we select people randomly from a base population, we are in charge of the sampling and can do so in an independent manner. Since the distribution is the same, this is a case of i.i.d. random variables. Note that if the population is large, the distribution of a small number of draws actually behaves similar to an i.i.d. draw, even if we sample without replacement.

Words in text documents are not independent because they follow certain compositional rules. For example, it is likely to find a noun preceded by an article.

If a dice is rolled repeatedly, we consider each roll an independent draw from the same distribution, hence this is an iid process.

Temperature measurements are highly correlated in time, although winter in Boston, where MIT is can sometimes make you think otherwise. Roughly speaking, if Monday has a warm weather, you would probably not expect Tuesday to be freezing cold.

I.I.D. assumption

1/1 point (graded)

What happens with our assumption of the observations of the kiss orientation being i.i.d. Bernoulli if we assume that the preferred orientation changes with the time of day?

- ○ The observations will always be dependent, so it is violated
- ● We will have to be more careful about how we collect observations
- ○ No matter how we sample, we will still have i.i.d. observations

✔

Solution:

The model will be more complicated, in particular, we cannot simply assume that all observations are $\mathrm{Ber}(p)$ across different times of the day. However, when we draw our collection times randomly from the hours of the day, we could still assume that there is some p' that corresponds to the average of personal preferences throughout the day, and still work under the i.i.d. assumption.

Lecture 2 - Probability Redux (refresher)

1. Objectives

Bookmark this page

1. Recall the statements of the **(strong/weak) law of large numbers** and the **central limit theorem** and know to apply these for large sample sizes.
2. Apply **Hoeffding's inequality** to the sample means of bounded i.i.d. random variables.
3. Recall the probability density function and properties of the **Gaussian distribution** .
4. Use **Gaussian probability tables** to obtain probabilities and **quantiles** .
5. Distinguish between **convergence almost surely** , **convergence in probability** and **convergence in distribution** , understand that these notions are from strongest to weakest.
6. Determine convergence of sums and products of sequences that converge almost surely or in probability.
7. Apply **Slutsky's theorem** to the sum and product of a sequence that converges in distribution and and another that converges in probability to a constant.
8. Use the **continuous mapping theorem** to determine convergence of sequences of a function of random variables.

Averages of random variables: Laws of Large Numbers and Central Limit Theorem

Let $X, X_1, X_2, \ldots, X_n$ be i.i.d. random variables, with $\mu = \mathbb{E}[X]$ and $\sigma^2 = \mathsf{Var}[X]$.

- Laws (weak and strong) of large numbers (LLN):

$$\bar{X}_n := \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow[n\to\infty]{\mathbf{P},\ \text{a.s.}} \mu$$

  where the convergence is in probability (as denoted by $\mathbf{P}$ on the convergence arrow) and almost surely (as denoted by a.s. on the arrow) for the weak and strong laws respectively.

- Central limit theorem (CLT):

$$\sqrt{n}\,\frac{\bar{X}_n - \mu}{\sigma} \xrightarrow[n\to\infty]{(d)} \mathcal{N}(0,1)$$

$$\text{or equivalently,}\quad \sqrt{n}\,(\bar{X}_n - \mu) \xrightarrow[n\to\infty]{(d)} \mathcal{N}(0,\sigma^2)$$

  where the convergence is in distribution, as denoted by (d) on top of the convergence arrow.

We will revisit the different modes of convergence near the end of this lecture.

Note : In *6.431x: Probability–the Science of Uncertainty and Data*, we used yet another equivalent formulation of the CLT:

$$\frac{S_n - n\mu}{\sqrt{n}\sigma} \xrightarrow[n\to\infty]{(d)} \mathcal{N}(0,1)$$

where $S_n = \sum_{i=1}^{n} X_i$ is the sum (not the average) of X_i.

Let $X_1, X_2, \ldots, X_n$ be i.i.d. **standard normal random variables**. For a finite n, what is the distribution of

$$\overline{X}_n = \frac{X_1 + X_2 + \cdots + X_n}{n}?$$

- (selected) A Gaussian.
- A χ^2-distribution.
- Cannot be determined for finite n, but asymptotically Gaussian.

✔

In terms of n, what are the variance and mean of $\overline{X}_n$?

$\mathsf{Var}(\overline{X}_n) =$ 1/n ✔ **Answer:** 1/n

$\frac{1}{n}$

$\mathbb{E}[\overline{X}_n] =$ 0 ✔ **Answer:** 0

0

Solution:

Since the sum of i.i.d. Gaussian random variables is also Gaussian, we deduce first that $X_1 + \cdots + X_n \sim N(0, n)$. Multiplying by $1/n$, we get $\overline{X}_n \sim N(0, 1/n)$, as scaling a random variable with a constant c scales its variance by c^2.

Therefore, $\overline{X}_n$ is a Gaussian random variable with mean 0 and variance $1/n$.

CLT Concept Check

0/1 point (graded)

Let $X_1, X_2, \ldots, X_n$ be a sequence of i.i.d. random variables with $\mathbb{E}[X] = \mu$, and $\mathsf{Var}(X) = \sigma^2$. Assuming that n is very large, according to the Central Limit Theorem, what is the best approximate characterization of the distribution of $\overline{X}_n$?

- ○ $N(0, 1)$.
- ○ $N(\mu, \sigma^2/n)$. ✔
- ● $N(0, \sigma^2/n)$.
- ○ Depends on the distribution of X.

✖

Solution:

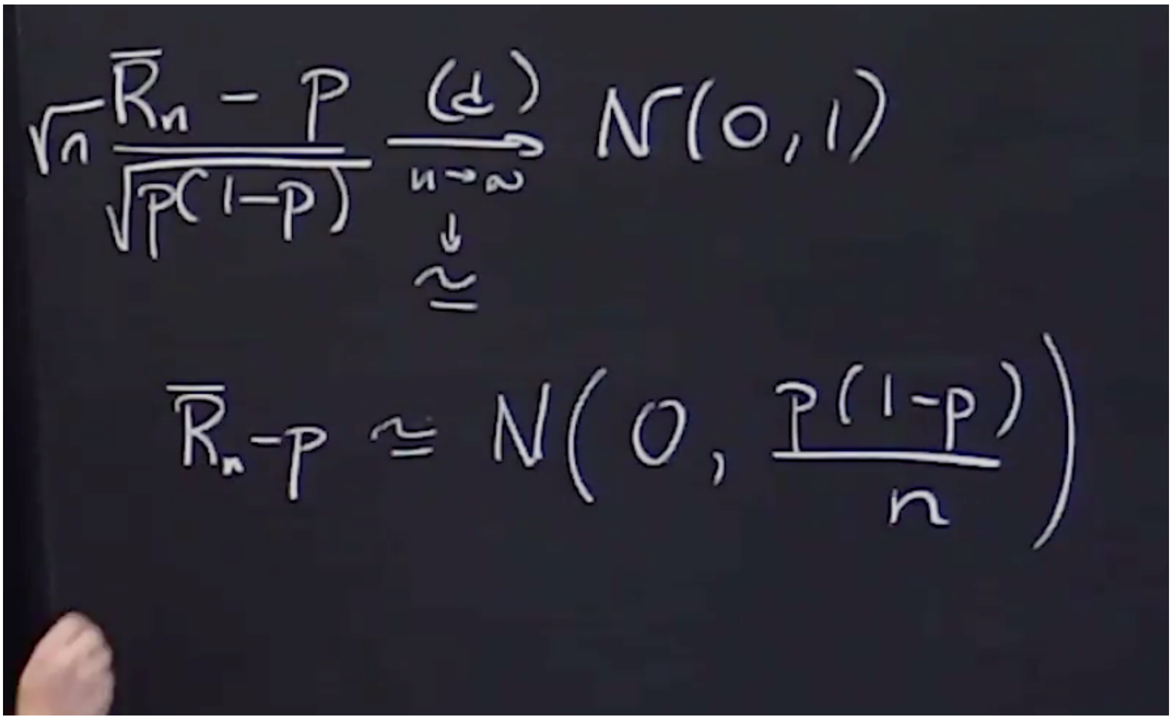
The correct choice is the second choice. We know by the Central Limit Theorem that

$$\frac{\sqrt{n}\,(\overline{X}_n - \mu)}{\sigma} \xrightarrow[n\to\infty]{d} \mathcal{N}(0, 1)$$

in distribution. Therefore, we can use approximate normality (as stated in the problem preamble), and get

$$\overline{X}_n \approx N(\mu, \sigma^2/n)\,.$$

Hoeffding's Inequality



Recall from the video the **Hoeffding's Inequality** :

Given n ($n > 0$) i.i.d. random variables $X_1, X_2, \ldots, X_n \overset{iid}{\sim} X$ that are almost surely **bounded** – meaning $\mathbf{P}(X \notin [a,b]) = 0$.

$$\mathbf{P}\left(\left|\overline{X}_n - \mathbb{E}[X]\right| \geq \epsilon\right) \leq 2\exp\left(-\frac{2n\epsilon^2}{(b-a)^2}\right) \qquad \text{for all } \epsilon > 0.$$

Unlike for the central limit theorem, here the **sample size n does not need to be large**.

Hoeffding's Inequality practice

0/1 point (graded)

Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \text{Unif}(0,b)$ be n i.i.d. uniform random variables on the interval $[0,b]$ for some positive b.

Using Hoeffding's inequality, which of the following can you conclude to be true? (Choose all that apply.)

- [x] $\mathbf{P}\left(\left|\overline{X}_n - \frac{b}{2}\right| \geq \frac{c}{n}\right) \leq 2e^{\frac{-2c^2}{b^2}}$ for $n = 3$
- [x] $\mathbf{P}\left(\left|\overline{X}_n - \frac{b}{2}\right| \geq \frac{c}{n}\right) \leq 2e^{\frac{-2c^2}{b^2}}$ for $n = 300$
- [x] $\mathbf{P}\left(\left|\overline{X}_n - \frac{b}{2}\right| \geq \frac{c}{\sqrt{n}}\right) \leq 2e^{\frac{-2c^2}{b^2}}$ for $n = 5$ ✔
- [x] $\mathbf{P}\left(\left|\overline{X}_n - \frac{b}{2}\right| \geq \frac{c}{\sqrt{n}}\right) \leq 2e^{\frac{-2c^2}{b^2}}$ for $n = 10$ ✔
- [x] $\mathbf{P}\left(\left|\overline{X}_n - \frac{b}{2}\right| \geq c\right) \leq 2e^{\frac{-2c^2}{b^2}}$ for $n = 10$ ✔
- [x] $\mathbf{P}\left(\left|\overline{X}_n - \frac{b}{2}\right| \geq c\right) \leq 2e^{\frac{-2c^2}{b^2}}$ for $n = 10000$ ✔

Solution:

Given that the X_i's are uniform and hence bounded, Hoeffding inequality holds, with mean $\mathbb{E}[X] = \frac{b}{2}$, and for any positive sample size n.

$$\mathbf{P}\left(\left|\overline{X}_n - \frac{b}{2}\right| \geq \epsilon\right) \leq 2e^{-\frac{2n\epsilon^2}{b^2}} \qquad \text{for all } \epsilon > 0.$$

The different answer choices involve different expressions for ϵ and different values of n, but since $n > 0$ in all choices, we only need to consider the effects of the ϵ.
In all choices, $\epsilon = \dfrac{c}{n^k}$: $k = 1$ in the first two choices, $k = 1/2$ in the third and fourth choices, and $k = 0$ in the last two choices. Plugging the expression for ϵ into Hoeffding's inequality, we have

$$\begin{aligned}\mathbf{P}\left(\left|\overline{X}_n - \frac{b}{2}\right| \geq \frac{c}{n^k}\right) &\leq 2e^{-\frac{2n}{b^2}\frac{c^2}{n^{2k}}}\\ &= 2e^{-\frac{2c^2}{b^2 n^{2k-1}}} \leq 2e^{-\frac{2c^2}{b^2}} \qquad \text{for } 2k-1 \leq 0.\end{aligned}$$

Since $2k - 1 \leq 0$ in the last four choices, that is, $\epsilon = \dfrac{c}{n^k}$ for $k \leq 1/2$, the probabilities in these choices are bounded above by the given quantity $2e^{-\frac{2c^2}{b^2}}$.
Remark: The Hoeffding equality holds for any positive n, even when n is small, including the extreme case $n = 1$.

Probability review: Markov and Chebyshev inequalities

Recall that in Unit 8 of the course *6.431x Probability–the Science of Uncertainty and Data*, we have seen two other inequalities which are upper bounds on $\mathbf{P}(X \geq t)$ based on the mean and variance of X.

Markov inequality

For a random variable $X \geq 0$ with mean $\mu > 0$, and any number $t > 0$:

$$\mathbf{P}(X \geq t) \leq \frac{\mu}{t}.$$

Note that the Markov inequality is restricted to **non-negative** random variables.

Chebyshev inequality

For a random variable X with (finite) mean μ and variance σ^2, and for any number $t \geq 0$,

$$\mathbf{P}(|X-\mu| \geq t) \leq \frac{\sigma^2}{t^2}.$$

Remark:

When Markov inequality is applied to $(X-\mu)^2$, we obtain Chebyshev's inequality. Markov inequality is also used in the proof of Hoeffding's inequality.

Hoeffding versus Chebyshev

4/4 points (graded)

Let $X_1, X_2, \ldots, X_n \stackrel{iid}{\sim} \mathsf{Unif}(0, b)$ be n i.i.d. uniform random variables on the interval $[0, b]$ for some positive b. Suppose n is small (i.e. $n < 30$) so that the central limit theorem is not justified.

Find an upper bound on the probability that the sample mean is "far away" from the expectation (the true mean) of X. More specifically, find the respective upper bounds given by the Chebyshev and Hoeffding inequalities on the following probability:

$$\mathbf{P}\left(\left|\overline{X}_n - \mathbb{E}[X]\right| \geq c\frac{\sigma}{\sqrt{n}}\right) \qquad \text{where } \sigma^2 = \mathsf{Var} X_i$$

for $c = 2$ and $c = 6$.

Hint: Each answer is numerical.

Using **Chebyshev** inequality:

$\mathbf{P}\left(\left|\overline{X}_n - \mathbb{E}[X]\right| \geq 2\frac{\sigma}{\sqrt{n}}\right) \leq$ 1/4 ✔ **Answer:** 1/4

$\mathbf{P}\left(\left|\overline{X}_n - \mathbb{E}[X]\right| \geq 6\frac{\sigma}{\sqrt{n}}\right) \leq$ 1/36 ✔ **Answer:** 1/36

Using **Hoeffding** inequality:

$\mathbf{P}\left(\left|\overline{X}_n - \mathbb{E}[X]\right| \geq 2\frac{\sigma}{\sqrt{n}}\right) \leq$ 1.0268 ✔ **Answer:** 2*e^(-2/3)

$\mathbf{P}\left(\left|\overline{X}_n - \mathbb{E}[X]\right| \geq 6\frac{\sigma}{\sqrt{n}}\right) \leq$ 0.0050 ✔ **Answer:** 2*e^(-6)

Solution:

Chebyshev: Since the variance of $\overline{X}_n$ is $\frac{\sigma^2}{n}$, Chebyshev inequality gives

$$\mathbf{P}\left(\left|\overline{X}_n - \mathbb{E}[X]\right| \geq t\right) \leq \frac{\sigma^2/n}{t^2}$$

Substitute $t = c\frac{\sigma}{\sqrt{n}}$, we have

$$\mathbf{P}\left(\left|\overline{X}_n - \mathbb{E}[X]\right| \geq c\frac{\sigma}{\sqrt{n}}\right) \leq \frac{1}{c^2}.$$

Hoeffding: On the other hand, substituting $\epsilon = c\frac{\sigma}{\sqrt{n}}$ in Hoeffding's inequality, we have

$$\begin{aligned}\mathbf{P}\left(\left|\overline{X}_n - \mathbb{E}[X]\right| \geq c\frac{\sigma}{\sqrt{n}}\right) &\leq 2\exp\left(-2c^2\frac{\sigma^2}{b^2}\right)\\ &\leq 2\exp\left(-2c^2\frac{1}{12}\right) = 2\exp\left(-\frac{c^2}{6}\right) \qquad \text{since } \sigma^2 = \frac{b^2}{12} \text{ for } X_i \sim \text{Unif}(0,b).\end{aligned}$$

Numerical bounds: Finally, plug in $c = 2, 6$ to get the following numerical upper bounds:

	$c=2$	$c=6$
Chebyshev:	$1/4 = 0.25$	$1/36 = 0.0278$
Hoeffding:	$2e^{-4/6} = 1.027$	$2e^{-36/6} = 0.00496$

Remark: When c is small, Chebyshev may give a better bound. But as c increases, the bound given by Hoeffding decays exponentially in c^2 while the bound given by Chebysheve decays only by $\frac{1}{c^2}$.

They are variances of different random variables.

$\text{Var}(X_i) \neq Var(\overline{X}_n)$

Chebyshev inequality says that:

$$P(|X - E[X]| \geqslant t) \leqslant \frac{\text{Var}(X)}{t^2}$$

In exercise, we are asked to compute the bounds for the sample mean, so in place of X we have $\overline{X}_n$.

$$\overline{X}_n = \frac{X_1 + ... + X_n}{n}$$

$$\text{Var}(\overline{X}_n) = \text{Var}\left(\frac{X_1 + ... + X_n}{n}\right) = \frac{1}{n^2}\cdot\text{Var}(X_1 + ... + X_n) = \frac{1}{n^2}\cdot n\cdot\text{Var}(X_i) = \frac{\sigma^2}{n}$$

Hoeffding's inequality says that:

$$\mathbf{P}\left(\left|\overline{X}_n - \mathbb{E}[X]\right| \geq \epsilon\right) \leq 2\exp\left(-\frac{2n\epsilon^2}{(b-a)^2}\right) \qquad \text{for all } \epsilon > 0.$$

In place of ϵ we have $c\frac{\sigma}{\sqrt{n}}$, and we are told in the problem desription that $\sigma^2 = \text{Var}(X_i)$, which is $\text{Unif}(0, b)$.

Probability review: PDF of Gaussian distribution

2/2 points (graded)

In practice, it is not often that you will need to work directly with the probability density function (pdf) of Gaussian variables. Nonetheless, we will make sure we know how to manipulate the (pdf) in the next two problems.

If the pdf X is a Gaussian variable is

$$f_X(x) = \frac{n}{3\sqrt{2\pi}}\exp\left(-\frac{n^2(x-2)^2}{18}\right),$$

then what is the mean μ and variance σ^2 of X?
(Enter your answer in terms of n.)

$\mu =$ 2 ✔ **Answer:** 2

2

$\sigma^2 =$ 9/n^2 ✔ **Answer:** 9/n^2

$\frac{9}{n^2}$

Solution:

Comparing

$$f_X(x) = \frac{n}{3\sqrt{2\pi}}\exp\left(-\frac{n^2(x-2)^2}{18}\right) = \frac{1}{(3/n)\sqrt{2\pi}}\exp\left(-\frac{(x-2)^2}{2(3/n)^2}\right)$$

with

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

yields $\mu = 2$ and $\sigma^2 = \frac{9}{n^2}$.

Probability review: PDF of Gaussian distribution

0/1 point (graded)

Let $X \sim \mathcal{N}(\mu, \sigma^2)$, i.e. the pdf of X is

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

Let $Y = 2X$. Write down the pdf of the random variable Y. (Your answer should be in terms of y, σ and μ. Type **mu** for μ, **sigma** for σ.)

$f_Y(y) =$ 1/(sigma*sqrt(2*pi)) * (✘ **Answer:** 1/(2*sigma*sqrt(2*pi))* exp(-(y-2*mu)^2/(8*sigma^2))

$\frac{1}{\sigma\cdot\sqrt{2\cdot\pi}}\cdot\exp\left(\frac{(y-2\cdot\mu)^2}{8\cdot\sigma^2}\right)$

Solution:

If $X \sim \mathcal{N}\left(\mu, \sigma^2\right)$, then $Y = 2X \sim \mathcal{N}\left(2\mu, 4\sigma^2\right)$ by the following general properties of expectations and variance:

$$\begin{aligned} \mathbb{E}\left[2X\right] &= 2\mathbb{E}\left[X\right] \\ \mathsf{Var}\left[2X\right] &= 2^2\mathsf{Var}\left[X\right] = 4\mathsf{Var}\left[X\right]. \end{aligned}$$

Therefore,

$$f_Y\left(y\right) = \frac{1}{2\sigma\sqrt{2\pi}}\exp\left(-\frac{\left(y-2\mu\right)^2}{2\left(4\sigma^2\right)}\right).$$

Alternate solution: In general, for any continuous random variables X and any continuous monotonous (i.e. always increasing or always decreasing) function g, such that $Y = g\left(X\right)$, the pdf of Y is given by:

$$f_Y\left(y\right) = \frac{f_X\left(x\right)}{\left|g'\left(x\right)\right|} \qquad \text{where } x = g^{-1}\left(y\right).$$

In this problem, $X \sim \mathcal{N}\left(\mu, \sigma^2\right)$, $Y = g\left(X\right) = 2X$, and $g'\left(x\right) = 2$. Therefore:

$$\begin{aligned} f_Y\left(y\right) &= \frac{f_X\left(\frac{y}{2}\right)}{\left|g'\left(\frac{y}{2}\right)\right|} \\ &= \frac{1}{g'\left(y/2\right)\sigma\sqrt{2\pi}}\exp\left(-\frac{\left(y/2-\mu\right)^2}{2\sigma^2}\right) \\ &= \frac{1}{2\sigma\sqrt{2\pi}}\exp\left(-\frac{\left(\left(y-2\mu\right)/2\right)^2}{2\sigma^2}\right) \\ &= \frac{1}{2\sigma\sqrt{2\pi}}\exp\left(-\frac{\left(\left(y-2\mu\right)\right)^2}{2\left(4\right)\sigma^2}\right) \end{aligned}$$

and we recover the same answer as above.

Standardization

1/1 point (graded)

Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables with mean μ and variance σ^2. Denote the sample mean by $\overline{X}_n = \frac{\sum_{i=1}^n X_i}{n}$.

Assume that n is large enough that the central limit theorem (clt) holds. Find a random variable Z with approximate distribution $\mathcal{N}\left(0,1\right)$, in terms of $\overline{X}_n$, n,μ and σ. (Note that μ and σ^2 refers to the mean and variance of X_i, not $\overline{X}_n$.)

(Type **barX_n** for $\overline{X}_n$, **mu** for μ and **sigma** for σ. Refer to the standard notation button below.)

$Z \sim \mathcal{N}\left(0,1\right)$ for $Z =$ ((barX_n - mu) / sigma)' ✔ **Answer:** (barX_n-mu)*sqrt(n)/sigma

$$\left(\frac{\overline{X}_n - \mu}{\sigma}\right)\cdot\sqrt{n}$$

Solution:

First, compute the mean and variance of $\overline{X}_n$:

$$\begin{aligned}\mathbb{E}\left[\overline{X}_n\right] &= \mathbb{E}\left[X_i\right] = \mu \\ \mathsf{Var}\left(\overline{X}_n\right) &= \frac{\sum_{i=1}^{n} \mathsf{Var}\left[X_i\right]}{n^2} = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}.\end{aligned}$$

Then, standardize by defining

$$\begin{aligned}Z &= \frac{\overline{X}_n - \mathbb{E}\left[\overline{X}_n\right]}{\sqrt{\mathsf{Var}\left(\overline{X}_n\right)}} \\ &= \frac{\overline{X}_n - \mu}{\sqrt{\sigma^2/n}} \\ &= \sqrt{n}\frac{\overline{X}_n - \mu}{\sigma}.\end{aligned}$$

By the clt, $Z \sim \mathcal{N}(0,1)$.

Transformation and Symmetry

0/1 point (graded)

Let $X \sim \mathcal{N}(2,2)$, i.e. X is a Gaussian variable with mean $\mu = 2$ and variance $\sigma^2 = 2$. Let $x > 0$.

Write $\mathbf{P}(X \geq -x)$ in terms of the cdf Φ of the **standard** Gaussian variable with a positive argument. In other words, your answer be in terms of of $\Phi(g(x))$, where $g(x)$ is a function of x which takes only **positive** values for $x > 0$.

(For example, if your answer is $1 + \Phi\left(\sqrt{5}x\right)$, type `1+Phi(sqrt(5)*x)`.)

$\mathbf{P}(X \geq -x) =$ 1- Phi(4+x) ✖ **Answer:** Phi((x+2)/sqrt(2))

$1 - \Phi(4 + x)$

Solution:

Standardizing $X \sim \mathcal{N}(2,2)$, we have $\frac{X-2}{\sqrt{2}} \sim \mathcal{N}(0,1)$. (The intuition here is that we are translating and re-scaling the density of X so that we end up with a standard Gaussian density.)

$$\begin{aligned}\mathbf{P}(X \geq -x) &= \mathbf{P}\left(\frac{X-2}{\sqrt{2}} \geq \frac{-x-2}{\sqrt{2}}\right) \\ &= \mathbf{P}\left(\frac{X-2}{\sqrt{2}} \leq \frac{x+2}{\sqrt{2}}\right) \quad \text{by symmetry} \\ &= \Phi\left(\frac{x+2}{\sqrt{2}}\right).\end{aligned}$$

The expression $1 - \Phi\left(\frac{-x-2}{\sqrt{2}}\right)$ gives the same value, but the argument is negative and so is not an accepted answer.

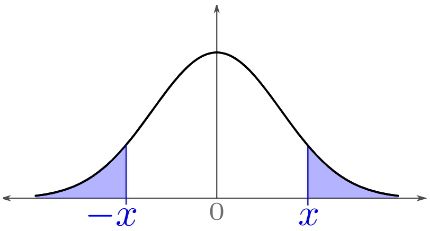
Remark: The symmetry is easiest to see by comparing the areas under the standard normal pdf corresponding to $\mathbf{P}(Z \geq -z)$ and $\mathbf{P}(Z \leq z)$ where $Z \sim \mathcal{N}(0,1)$ and $z > 0$.

Quantiles

1/1 point (graded)

The **quantile** of order $1-\alpha$ of a variable X, denoted by q_α (specific to a particular X), is the number such that $\mathbf{P}(X \leq q_\alpha) = 1-\alpha$.

Graphed below is the pdf of the normal distribution with generic/unknown (but fixed) variance σ^2. If the total area of the two shaded regions is 0.03, then what is x?
(Choose all that apply.)



The total area of the two shaded regions is 0.03.

- [ ] $\mathbf{P}(|X| \leq 0.03)$
- [ ] $\mathbf{P}(|X| \leq 0.015)$
- [ ] 0.97
- [ ] 0.985
- [ ] $q_{0.03}$
- [x] $q_{0.015}$

The total area of the two shaded regions equals $\mathbf{P}(|X| \geq x) = 0.03$. By symmetry, the probability in the positive tail is $\mathbf{P}(X \geq x) = 0.015$; hence $x = q_\alpha$ with $\alpha = 0.015$.
For the wrong choices:

- The first pair of choices mixed up the values of probability with the value of the variable.
- The third and fourth choices "0.97" and "0.985" are meant to play the role resembling $1-\alpha$ in this example, but these are wrong for the same reasons as the first pair of choices. In any case, to give a particular numerical value of x, the answer must depend on σ.
- The fifth choice would have been correct again if the area of one of the tails is 0.03.

Quantiles

1/1 point (graded)

Which of the following is the correct ordering of the numbers $q_{0.05}, 2q_{0.5}, q_{0.02}$, which are quantiles of a standard Gaussian variable?

- ○ $q_{0.02} < 2q_{0.5} < q_{0.05}$
- ● $2q_{0.5} < q_{0.05} < q_{0.02}$
- ○ $q_{0.05} < 2q_{0.5} < q_{0.02}$
- ○ $q_{0.05} < q_{0.02} < 2q_{0.5}$

✔

Solution:

Recall that q_α is the number such that $\mathbf{P}(X \geq q_\alpha) = \alpha$; that is, the probability of the tail to the right of q_α is α. Therefore, $q_{0.05} < q_{0.02}$. Since these are quantiles of a standard Gaussian variable, $q_{0.5} = 0$, and $2q_{0.5} = 0$. So the correct ordering is given by $2q_{0.5} < q_{0.05} < q_{0.02}$.
Remark:. In general, the quantiles of any continous random variable satisfes $q_a > q_b$ if $a < b$.

The score distribution of the final exam in a data science course follows a normal distribution with **mean** 70 and **standard deviation** 10.

Let α in $(0, 1)$. As a reminder, the quantile of order $1 - \alpha$ of a random variable X is the number q_α such that

$$\mathbf{P}(X \leq q_\alpha) = 1 - \alpha.$$

According to this distribution, what score do you need to get in order to be at the 90th percentile of the class, that is, in order that 90% of all students in the class have a score less than or equal to your score?

Use either the standard normal table below or any online calculator or software.

Normal Table

Show

(Do **Not** round to the integer.)

Required score: 82.9 ✔ **Answer:** 82.82

STANDARD NOTATION

Solution:

Given the final exam grade X follows a normal distribution with $\mu = 70$ and $\sigma = 10$, we can define a variable $Z = \frac{X-70}{10}$ that follows the standard normal distribution $\mathcal{N}(0, 1)$.

$$\mathbf{P}(X \leq t) = 0.9 \quad \text{if and only if} \quad \mathbf{P}\left(Z \leq \frac{t-70}{10}\right) = 0.9 \quad \text{if and only if} \quad \frac{t-70}{10} = q_{0.1} = \Phi^{-1}(0.9).$$

where $q_{0.1} = \Phi^{-1}$ is the inverse of the cdf of the standard normal distribution. Since $\Phi^{-1}(0.9) = 1.282$ (e.g. by using **qnorm(0.9)** in R), we have

$$t = 1.282 * 10 + 70 = 82.82.$$

Convergence in probability and in distribution 1

1.5/2 points (graded)

Let $(T_n)_{n\geq 1} = T_1, T_2, \ldots$ be a sequence of r.v.s such that

$$T_n \sim \text{Unif}\left(5 - \frac{1}{2n}, 5 + \frac{1}{2n}\right).$$

Given an arbitrary fixed number $0 < \delta < 1,$ find the smallest number N (in terms of δ) such that $\mathbf{P}(\{|T_n - 5| > \delta\}) = 0$ whenever $n > N$.

$N =$ delta-5 ✖ **Answer:** 1/(2*delta)

$\delta - 5$

Does $(T_n)_{n\geq 1}$ converge in probability to a constant? If so, what is the limiting value? Enter **DNE** if $(\{X_n\})$ does not converge in probability.

$(T_n)_{n\geq 1} \xrightarrow{\mathbf{P}}$ 5 ✔ **Answer:** 5

5

Does $(T_n)_{n\geq 1}$ converge in distribution?

- (selected) Yes
- No

✔

Let $F_n(t)$ be the cdf of T_n and $F(t)$ be the cdf of the constant limit. For which values of t does $\lim_{n\to\infty} F_n(t) = F(t)$? (Choose all that apply.)

- [x] $t < 5$
- [ ] $t = 5$
- [x] $t > 5$

✔

STANDARD NOTATION

Solution:

- Given a fixed $0 < \delta < 1$, since $T_n \sim \text{Unif}\left(5 - \frac{1}{2n}, 5 + \frac{1}{2n}\right)$, we know that $\mathbf{P}(|T_n - 5| > \delta) = 0$ whenever $\frac{1}{2n} < \delta$, or equivalently, for all $n > \frac{1}{2\delta}$.

- By the definition of convergence in probability, $T_n \xrightarrow[n\to\infty]{\mathbf{P}} 5$.

- Since convergence in probability implies convergence in distribution, $T_n \xrightarrow[n\to\infty]{\text{d.}} 5$.

- The cdf F_n of T_n is a piecewise linear function with value 0 for all $t \leq 5 - \frac{1}{2n}$, 1 for all $t \geq 5 + \frac{1}{2n}$, and a line connecting the points $(t, F_n(t)) = \left(5 - \frac{1}{2n}, 0\right)$ and $(t, F_n(t)) = \left(5 + \frac{1}{2n}, 1\right)$ in the interval $5 - \frac{1}{2n} \leq t \leq 5 + \frac{1}{2n}$. In particular, $F_n(5) = \frac{1}{2}$ for all n. On the other hand, the cdf F of the contant 5 is $F(t) = 0$ when $t < 5$, and $F(t) = 1$ when $t \geq 5$. Therefore, $F_n(t) \xrightarrow[n\to\infty]{} F(t)$ for all $t \neq 5$.
  Remark: We have just verified that T_n indeed converges in distribution to the deterministic limit 5, that is, $F_n(t) \xrightarrow[n\to\infty]{} F(t)$ for all t where $F(t)$ is continuous.

Convergence in probability and in distribution 2

3/4 points (graded)

Let $(Y_n)_{n\geq 1}$ be a sequence of i.i.d. random variables with $Y_n \sim \text{Unif}(0,1)$.

Let

$$M_n = \max(Y_1, Y_2, \ldots, Y_n).$$

For any fixed number $0 < \delta < 1$, find $\mathbf{P}(|M_n - 1| > \delta\})$. (Type **delta** for δ.)

$\mathbf{P}(|M_n - 1| > \delta\}) =$ delta+1 ✖ **Answer:** (1-delta)^n

$\delta + 1$

Does the sequence $(M_n)_{n\geq 1}$ converge in probability to a constant? If yes, enter the value of the constant limit; if no, enter **DNE**.

$(M_n)_{n\geq 1} \xrightarrow{\mathbf{P}}$ 1 ✔ **Answer:** 1

1

Find the CDF $F_{M_n}(x)$ for $0 \leq x \leq 1$.

$F_{M_n}(x) = P(M_n \leq x) =$ x^n ✔ **Answer:** x^n

x^n

Does $(M_n)_{n\geq 1}$ converge in distribution?

- (●) Yes
- ( ) No

Solution:

Note that M_n is always at most one, so $|M_n - 1| \geq \delta$ can be replaced with $1 - M_n \geq \delta$.

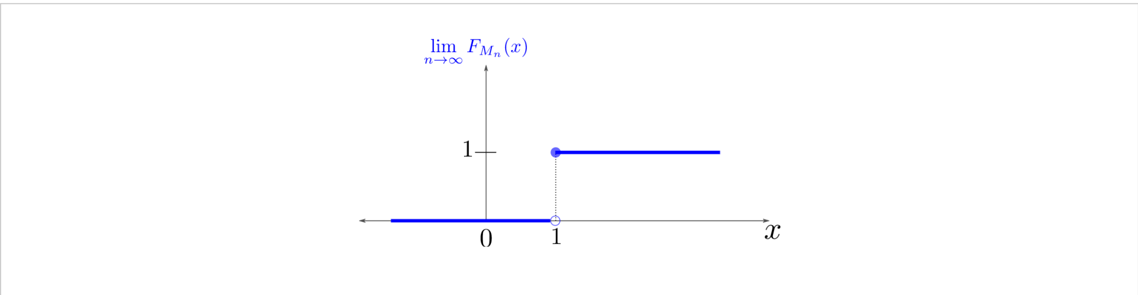
$$\begin{aligned} \mathbf{P}(|M_n - 1| \geq \delta) &= \mathbf{P}(1 - M_n \geq \delta) \\ &= \mathbf{P}(M_n \leq 1 - \delta) \\ &= \mathbf{P}(Y_1 \leq 1 - \delta)\,\mathbf{P}(Y_2 \leq 1 - \delta) \cdots \mathbf{P}(Y_n \leq 1 - \delta) \qquad \text{since } Y_i \text{ independent} \\ &= (1 - \delta)^n \xrightarrow[n\to\infty]{} 0 \qquad \text{since } 0 < (1 - \delta) < 1. \end{aligned}$$

Hence, the sequence $(M_n)_{n\geq 1}$ converges in probability to the deterministic limit $M = 1$. This implies that is also converges in distribution to the same limit.

The CDF is computed in the same way:

$$F_{M_n}(x) = \mathbf{P}(M_n \leq x) = \mathbf{P}(Y_1 \leq x) \cdots \mathbf{P}(Y_n \leq x) = x^n \qquad \text{for } 0 \leq x \leq 1.$$

We already know that $(M_n)_{n\geq 1}$ converges in distribution to M; here we check directly through definition. As $n \to \infty$, $F_{M_n}(x)$ approaches the step function shown below



which coincides with the CDF of the limit M we found above; hence, indeed $M_n \xrightarrow[n\to\infty]{(d.)} M$.

Expectations and convergence in probability

3/3 points (graded)

Let $(T_n)_{n\geq 1}$ be a sequence of r.v.s such that for each n, T_n takes only two possible values 0 and 2^n with the following probabilities:

$$\begin{aligned} \mathbf{P}(T_n = 0) &= 1 - \frac{1}{n} \\ \mathbf{P}\left(T_n = 2^n\right) &= \frac{1}{n}. \end{aligned}$$

Does the sequence $(T_n)_{n\geq 1}$ converge in probability to a constant? If so, enter the limiting value; if not, enter **DNE**.

$T_n \xrightarrow{\mathbf{P}}$ 0 ✔

0

Compute $\mathbb{E}[T_n]$ in terms of n.

$\mathbb{E}[T_n] =$ 2^n / n ✔

$\frac{2^n}{n}$

Does the sequence of expectations $\mathbb{E}[T_n]$ converge? If so, enter the limiting value; if not, enter **DNE**.

$\lim_{n\to\infty} \mathbb{E}[T_n] =$ DNE ✔

DNE

Solution:

For any $\epsilon > 0$,

$$\mathbf{P}(|T_n - 0| > \epsilon) = \frac{1}{n} \longrightarrow 0.$$

Therefore, $(T_i)_{n\geq 1}$ converges in probability to the deterministic limit 0.
However,

$$\mathbb{E}[T_n] = \frac{2^n}{n} \longrightarrow \infty.$$

Hence, the sequence $(\mathbb{E}[T_n])_{n\geq 1}$ does not converge.
Remark: Convergence in probability does not imply convergence of expectation values.

Probability review: the (Strong) Law of Large Numbers

1/1 point (graded)

A digital signal receiver decodes bits of incoming signal as 0s or 1 and makes an error in decoding a bit with probability 10^{-4}.

Assuming decoding success is independent for different bits, as the receiver receives more and more signals, what is the fraction of erroneously decoded bits?

Fraction of errors: 1/10000 ✔ **Answer:** 10^(-4)

Solution:

The transmission of each bit of data can be modelled as independent Bernoulli random variable X_i with expectation of error $\mathbb{E}[X_i] = 10^{-4}$.
The (strong/weak) law of large number states that

$$\frac{\sum_{i=1}^n X_i}{n} \xrightarrow[n\to\infty]{\text{(a.s./P)}} \mathbb{E}[X].$$

In particular, the strong law says that **with probability 1**, the fraction of errors approach 10^{-4}.

Addition, Multiplication, Division for Convergence almost surely and in probability :

Addition, Multiplication, and Division preserves convergence almost surely (a.s.) and convergence in probability ($\mathbf{P}$).

More precisely, assume

$$T_n \xrightarrow[n\to\infty]{\text{a.s.}/\mathbf{P}} T \qquad \text{and} \qquad U_n \xrightarrow[n\to\infty]{\text{a.s.}/\mathbf{P}} U$$

Then,

- $T_n + U_n \xrightarrow[n\to\infty]{\text{a.s.}/\mathbf{P}} T + U,$
- $T_n U_n \xrightarrow[n\to\infty]{\text{a.s.}/\mathbf{P}} TU,$
- If in addition, $U \neq 0$ a.s., then $\dfrac{T_n}{U_n} \xrightarrow[n\to\infty]{\text{a.s.}/\mathbf{P}} \dfrac{T}{U}.$

Warning: In general, these rules **do not** apply to convergence in distribution (d).

Slutsky's Theorem

For convergence in distribution, the Slutsky's Theorem will be our main tool.

Let $(T_n)\,,(U_n)$ be two sequences of r.v., such that:

- $T_n \xrightarrow[n\to\infty]{(d)} T$
- $U_n \xrightarrow[n\to\infty]{\mathbf{P}} u$

where T is a r.v. and u is a given real number (deterministic limit: $\mathbf{P}\,(U = u) = 1$). Then,

- $T_n + U_n \xrightarrow[n\to\infty]{(d)} T + u,$
- $T_n U_n \xrightarrow[n\to\infty]{(d)} Tu,$
- If in addition, $u \neq 0$, then $\dfrac{T_n}{U_n} \xrightarrow[n\to\infty]{(d)} \dfrac{T}{u}.$

Continuous Mapping Theorem

If f is a continuous function:

$$T_n \xrightarrow[n\to\infty]{\text{a.s.}/\mathbf{P}/(d)} T \quad \Rightarrow \quad f\,(T_n) \xrightarrow[n\to\infty]{} f\,(T)\,.$$

Convergence in distribution

4/4 points (graded)

Let X_n be a sequence of random variables that are converging **in probability** to another random variable X. Let Y_n be a sequence of random variables that are converging **in probability** to another random variable Y.

For each of the statements below, choose true ("This statement is always true") or false ("This statement is sometimes false"). Keep in mind that "convergence in probability" is stronger than "convergence in distribution".

- $X_n + Y_n \longrightarrow X + Y$ in distribution.

  - (selected) True
  - False

- $X_n Y_n \longrightarrow XY$ in distribution.

  - (selected) True
  - False

- $X_n / Y_n \longrightarrow X/Y$ in distribution, provided Y is constant.

  - True
  - (selected) False

- $X_n^2 - 2X_n + 5 \longrightarrow X^2 - 2X + 5$ in distribution.

  - (selected) True
  - False

Solution:

- True. Sums of sequences that converge in probability converge in probability, and convergence in probability implies convergence in distribution.
- True. Since both X_n and Y_n converge in probability to X and Y respectively, $X_n Y_n$ converges in probability, and hence in distribution, to XY.
- False. Even though Y_n converges to a constant, this constant can very well be 0, in which case we do not have the desired convergence.
- True. This is a consequence of continuous mapping theorem, since the function $g(x) = x^2 - 2x + 5$ is continuous, $X_n \longrightarrow X$ in distribution implies $g(X_n) \longrightarrow g(X)$ in distribution.

Applying Slutsky's and the Continuous Mapping theorems

0/1 point (graded)

Given the following:

- $Z_1, Z_2, \ldots, Z_n, \ldots$ is a sequence of random variables that converge in distribution to another random variable Z;
- $Y_1, Y_2, \ldots, Y_n, \ldots$ is a sequence of random variables each of which takes value in the interval $(0, 1)$, and which converges in probability to a constant c in $(0, 1)$;
- $f(x) = \sqrt{x(1-x)}$.

Does $Z_n \dfrac{f(Y_n)}{f(c)}$ converge in distribution? If yes, enter the limit in terms of Z, Y and c; if no, enter **DNE**.

$Z_n \dfrac{f(Y_n)}{f(c)} \xrightarrow{\text{d}}$ Z*(Y/c) ✖ **Answer:** Z

$Z \cdot \left(\frac{Y}{c}\right)$

Solution:

Since f is continuous in $(0, 1)$, $f(Y_n)$ converges in probability to $f(c)$ by the continous mapping theorem. Since $f(c)$ is a constant, we have $\frac{f(Y_n)}{f(c)}$ converges in probability to 1. Finally, since Z_n converges in distribution to Z and $\frac{f(Y_n)}{f(c)}$ converges in probability to a constant, by Slutsky's Theorem, $Z_n \dfrac{f(Y_n)}{f(c)}$ converges in distribution to Z.