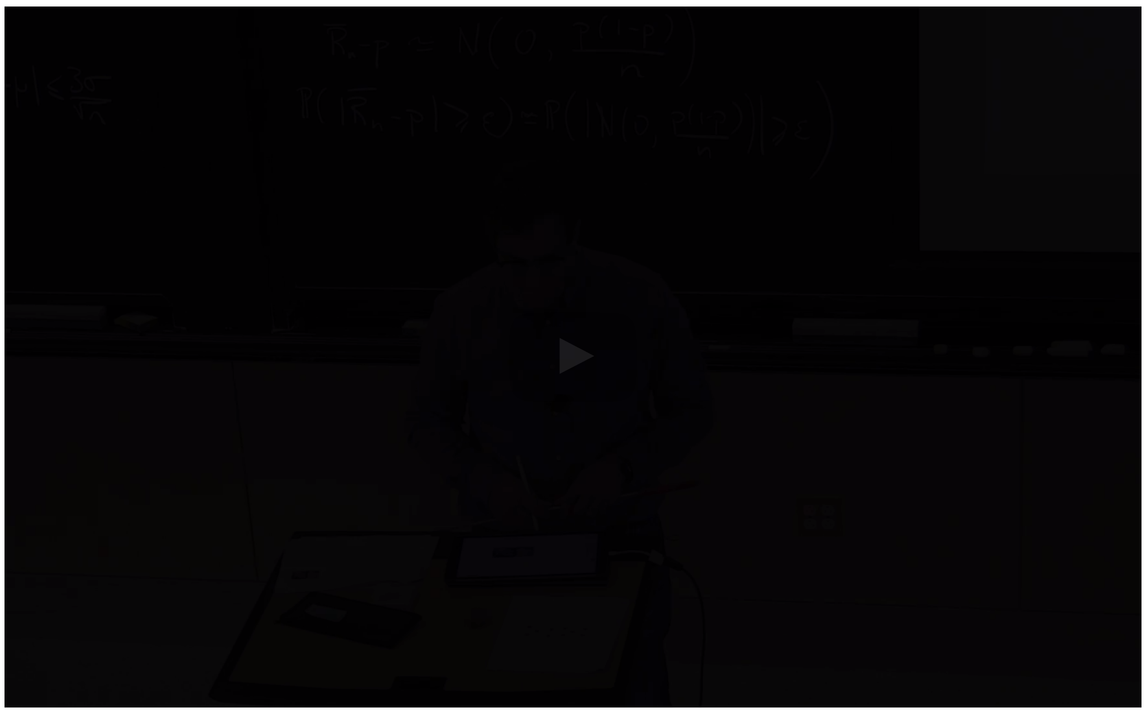
# 3. Hoeffding's Inequality

Bookmark this page

Exercises due May 25, 2021 19:59 EDT

**Small sample size of bounded random variables: Hoeffding's Inequality**



research, drug testing, all

this things--

they're not going to use have Hoeffding's inequality.

No one wants to do this because with the same amount of data,

you actually make less precise statements

than if you were to were use the central limit theorem.

So people prefer to rely on more precisely.

Any questions?

**So let's move on.**

17:15 / 17:15 1.50x

**Video**
Download video file

**Transcripts**
Download SubRip (.srt) file
Download Text (.txt) file

---

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

Given $n$ ($n > 0$) i.i.d. random variables $X_1, X_2, \ldots, X_n \stackrel{iid}{\sim} X$ 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

1/1 point (graded)

Let $X_1, X_2, \ldots, X_n \stackrel{iid}{\sim} \mathrm{Unif}(0, b)$ 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.)

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

Submit You have used 2 of 2 attempts

✔ Correct (1/1 point)

**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 $\mu$ and variance $\sigma^2$, and for any number $t > 0$,

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

**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 points possible (graded)
Let $X_1, X_2, \ldots, X_n \stackrel{iid}{\sim} \text{Unif}(0, b)$ 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)

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:

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

$\mathbf{P}\left(\left|\overline{X}_n - \mathbb{E}[X]\right| \geq 6\frac{\sigma}{\sqrt{n}}\right) \leq$

Using **Hoeffding** inequality:

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

$\mathbf{P}\left(\left|\overline{X}_n - \mathbb{E}[X]\right| \geq 6\frac{\sigma}{\sqrt{n}}\right) \leq$

Submit You have used 0 of 3 attempts

## Discussion

Hide Discussion

**Topic:** Unit 1 Introduction to statistics:Lecture 2: Probability Redux / 3. Hoeffding's Inequality

**Add a Post**

**All Posts**

### intuitive way to sovle Q1

discussion posted about 18 hours ago by vstone9

Q1 is a little bit tricky: while there are right answers directly calculated by the formula, others are not. But bear in mind, some answers provide more 'loose' range than the formula suggested. Then, choose them too because if the formula holds, these answers are right, i.e., even P(X'-E[X]>=' any thing')<=100% is a correct answer.

This post is visible to everyone.

0 responses

Preview

Submit

Previous Next

## edX

- About
- Affiliates
- edX for Business
- Open edX
- Careers
- News

## Legal

- Terms of Service & Honor Code
- Privacy Policy
- Accessibility Policy
- Trademark Policy
- Sitemap

## Connect

- Blog
- Contact Us
- Help Center
- Media Kit
- Donate

GET IT ON Google play    Download on the App Store