3. Hoeffding's Inequality

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

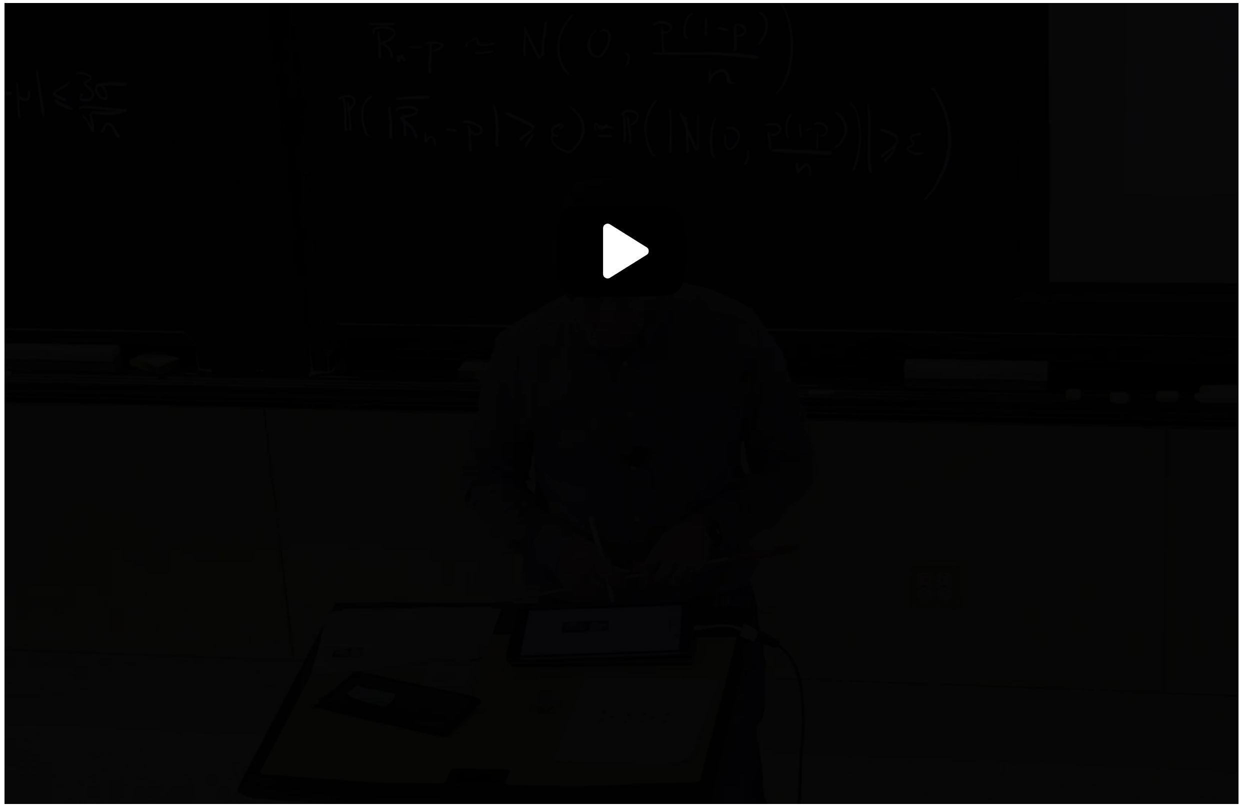
they're not going to use have Hoeffding's inequality.

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.

Generating Speech Output



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** :

Generating Speech Output

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 $\mathbf{P}\left(X \notin [a,b]\right) = 0$.

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

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 $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \mathsf{Unif}(0,b)$ 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 $n = 300$
- [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 $n = 5$ ✔

Generating Speech Output

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

✖

Solution:

Given that the X_i's are uniform and hence bounded, Hoeffding inequality holds, with mean $\mathbb{E}\left[X\right] = \frac{b}{2}$, 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

Generating Speech Output

$$\mathbf{P}\left(\left|\overline{X}_n - \frac{b}{2}\right| \geq \frac{c}{n^k}\right) \leq 2e^{-\frac{2n}{b^2}\frac{c^2}{n^{2k}}}$$

$$= 2e^{-\frac{2c^2}{b^2 n^{2k-1}}} \leq 2e^{-\frac{2c^2}{b^2}} \qquad \text{for } 2k-1 \leq 0.$$

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 $n = 1$.

Submit You have used 2 of 2 attempts

Answers are displayed within the problem

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 $t > 0$:

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

Generating Speech Output

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 $t \geq 0$,

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

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 $X_1, X_2, \ldots, X_n \overset{iid}{\sim} \mathsf{Unif}(0,b)$ 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:

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

Generating Speech Output $6.$

Hint: Each answer is numerical.

Using **Chebyshev** inequality:

$\mathbf{P}\left(\left|\overline{X}_n - \mathbb{E}\left[X\right]\right| \geq 2\frac{\sigma}{\sqrt{n}}\right) \leq$ 0.25 ✔ **Answer:** 1/4

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

Using **Hoeffding** inequality:

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

$\mathbf{P}\left(\left|\overline{X}_n - \mathbb{E}\left[X\right]\right| \geq 6\frac{\sigma}{\sqrt{n}}\right) \leq$ 0.004957504353332717 ✔ **Answer:** 2*e^(-6)

Solution:

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

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

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

Generating Speech Output

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

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

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

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 $\frac{1}{c^2}$.

Submit You have used 2 of 2 attempts

Answers are displayed within the problem

Generating Speech Output

Discussion

Hide Discussion

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

Add a Post

All Posts

The decimal of rounding and the value of e

discussion posted 4 days ago by anonymous

Both of these values are not specified and it makes me unconfidence with my answer.

I hope they are given in the question.

Thanks,

This post is visible to everyone.

Add a Response

1 response

navneethc

4 days ago - endorsed 3 days ago by **sudarsanvsr_mit** (Staff)

If you are referring to the problem of **Heoffding versus Chebyshev**, let me assure you that the answers come out to be nice fractions (where applicable) and that you can use e as it is, i.e. as suggested in the Standard Notation. I was green-ticked for both.

Generating Speech Output

Add a comment

Showing all responses

Add a response:

Preview

Submit

Learn About Verified Certificates

Generating Speech Output