Homework 1

Frederick J. Boehm

9/7/2020

Exercise 1

$$\begin{aligned}\mathbb{P}(|\hat{p}_n - p| \geq \epsilon) &\leq \frac{Var(\hat{p}_n)}{\epsilon^2}\\ &= \frac{p(1-p)}{n\epsilon^2}\end{aligned}$$

Exercise 2

We leverage the fact that the product $p(1-p)$ achieves a maximum value of $\frac{1}{4}$ for $p \in (0,1)$.

This leads us to the following bound:

$$\begin{aligned}\mathbb{P}(|\hat{p}_n - p| \geq \epsilon) &\leq \frac{Var(\hat{p}_n)}{\epsilon^2}\\ &= \frac{p(1-p)}{n\epsilon^2}\\ &\leq \frac{1}{4n\epsilon^2}\end{aligned}$$

Exercise 3

We set $n = 10,000$ and plug in the desired confidence level to get $\epsilon > 0$.

$$\begin{aligned}1 - 0.99 &= \mathbb{P}(|\hat{p}_n - p| \geq \epsilon)\\ &\leq \frac{1}{4n\epsilon^2}\\ &\implies \frac{1}{100} = \frac{1}{4 * 10^4\epsilon^2}\\ &\implies \epsilon^2 = \frac{1}{400}\\ &\implies \epsilon = \frac{1}{20}\end{aligned}$$

Exercise 4

We first consider $X \sim \text{Bernoulli}(p)$ and its cumulant-generating function.

$$\Lambda_X(\lambda) = \log \mathbb{E}[e^{\lambda X}]$$

We need to find $e^{\lambda X}$ for each value of X.

$$
\begin{aligned}
\mathbb{P}(X = 0) &= 1 - p \\
&= \mathbb{P}(e^{\lambda X} = 1) \\
&= 1 - \mathbb{P}(e^{\lambda X} = e^{\lambda})
\end{aligned}
$$

Hence, $\mathbb{E}(e^{\lambda X}) = 1 - p + pe^{\lambda}$.

We then have that the cgf is:

$$\Lambda_X(\lambda) = \log\{1 - p + pe^{\lambda}\}$$

We represent $S_n \sim \text{Binomial}(n, p)$ as a sum of n independent Bernoulli(p) random variables to see that

$$\Lambda_{S_n}(\lambda) = n \log\{1 - p + pe^{\lambda}\}$$

Exercise 5

For $X \sim \text{Bernoulli}(p)$, Chernov tells us that:

$$
\begin{aligned}
\mathbb{P}(X - p \geq \epsilon) &= \mathbb{P}(X \geq p + \epsilon) \\
&\leq \exp\left(-\sup_{\lambda \geq 0}\{\lambda(p+\epsilon) - \log(1 - p + pe^{\lambda})\}\right) \\
&= \exp\left(-\{\hat{\lambda}(p+\epsilon) - \log(1 - p + pe^{\hat{\lambda}})\}\right) \\
&= \exp\left(-(p+\epsilon)\log\left[\frac{(p+\epsilon)(1-p)}{p(1-p-\epsilon)}\right] - \log\left[1 - p + \frac{(p+\epsilon)(1-p)}{1-p-\epsilon}\right]\right) \\
&= \exp\left(\log\left[\frac{(p+\epsilon)^{p+\epsilon}(1-p)^{p+\epsilon}}{p^{p+\epsilon}(1-p-\epsilon)^{p+\epsilon}}\right] - \log\left[\frac{1-p}{1-p-\epsilon}\right]\right) \\
&= \frac{(p+\epsilon)^{p+\epsilon}(1-p)^{p+\epsilon-1}}{p^{p+\epsilon}(1-p-\epsilon)^{p+\epsilon-1}} \\
&= \frac{(p+\epsilon)^{p+\epsilon}}{p^{p+\epsilon}} \frac{(1-p)^{p+\epsilon-1}}{(1-p-\epsilon)^{p+\epsilon-1}} \\
&= \left(\frac{p}{p+\epsilon}\right)^{p+\epsilon}\left(\frac{1-p}{1-p-\epsilon}\right)^{1-p-\epsilon} \\
&= \exp\left(-R(p+\epsilon, p)\right)
\end{aligned}
$$

Note that we let $\hat{\lambda}$ denote the non-negative value of λ for which the maximum value of $\lambda(p + \epsilon) - \log\left(1 - p + pe^{\lambda}\right)$ is achieved.

Two steps remain. Below, we 1. establish the last equality from the above equations. We also 2. solve for $\hat{\lambda}$.

Establish the last equality

$$
\begin{aligned}
\exp\left(-R(p+\epsilon, p)\right) &= \exp\left(-(p+\epsilon)\log\left[\frac{p+\epsilon}{p}\right] - (1-p-\epsilon)\log\left[\frac{1-p-\epsilon}{1-p}\right]\right) \\
&= \left(\frac{p+\epsilon}{p}\right)^{-(p+\epsilon)}\left(\frac{1-p-\epsilon}{1-p}\right)^{-(1-p-\epsilon)}
\end{aligned}
$$

Solve for $\hat{\lambda}$

We differentiate the expression to be optimized with respect to λ.

$$\begin{aligned}
0 &= \frac{d}{d\lambda}\left[\lambda(p+\epsilon) - \log\left(1 - p - pe^{\lambda}\right)\right] \\
&= (p+\epsilon) - \frac{pe^{\lambda}}{1 - p - pe^{\lambda}} \\
&\implies (p+\epsilon)(1-p) = e^{\lambda}\left(p - p(p+\epsilon)\right) \\
&\implies e^{\hat{\lambda}} = \frac{(p+\epsilon)(1-p)}{p - p(p+\epsilon)}
\end{aligned}$$

From one Bernoulli RV to sum of iid Bernoulli RVs

We leverage the fact that the sum of n iid Bernoulli(p) random variables is a Binomial(n, p) random variable and the definition of the cumulant generating function (and resulting properties of cgf) to achieve the desired result.

Second bound

We need to establish the second bound, Equation 1.71 from the notes.

This follows from a sequence of calculations that is nearly identical to those above. First, we need the analog of 1.34 from 1.32.

$$\mathbb{P}(X \leq t) \leq \exp\left(-\sup_{\lambda \leq 0}\left[\lambda t - \Lambda_X(\lambda)\right]\right)$$

We then proceed in a manner like the above to get the result.

Exercise 6

$$\begin{aligned}
&g(\epsilon) = R(p+\epsilon, p) - 2\epsilon^2 \\
&\quad\implies \frac{d}{d\epsilon}g(\epsilon) = \frac{d}{d\epsilon}((p+\epsilon)\log\frac{p+\epsilon}{p} + (1-p-\epsilon)\log\frac{1-p-\epsilon}{1-p} - 2\epsilon^2 \\
&\quad= (p+\epsilon)\left(\frac{p}{p+\epsilon}\right)\frac{1}{p} + \log\frac{p+\epsilon}{p} + (1-p-\epsilon)\left(\frac{1-p}{1-p-\epsilon}\right)\left(\frac{-1}{1-p}\right) - \log\frac{1-p-\epsilon}{1-p} - 4\epsilon \\
&\quad\implies g'(0) = 0
\end{aligned}$$

Now, we calculate $g''(\xi)$.

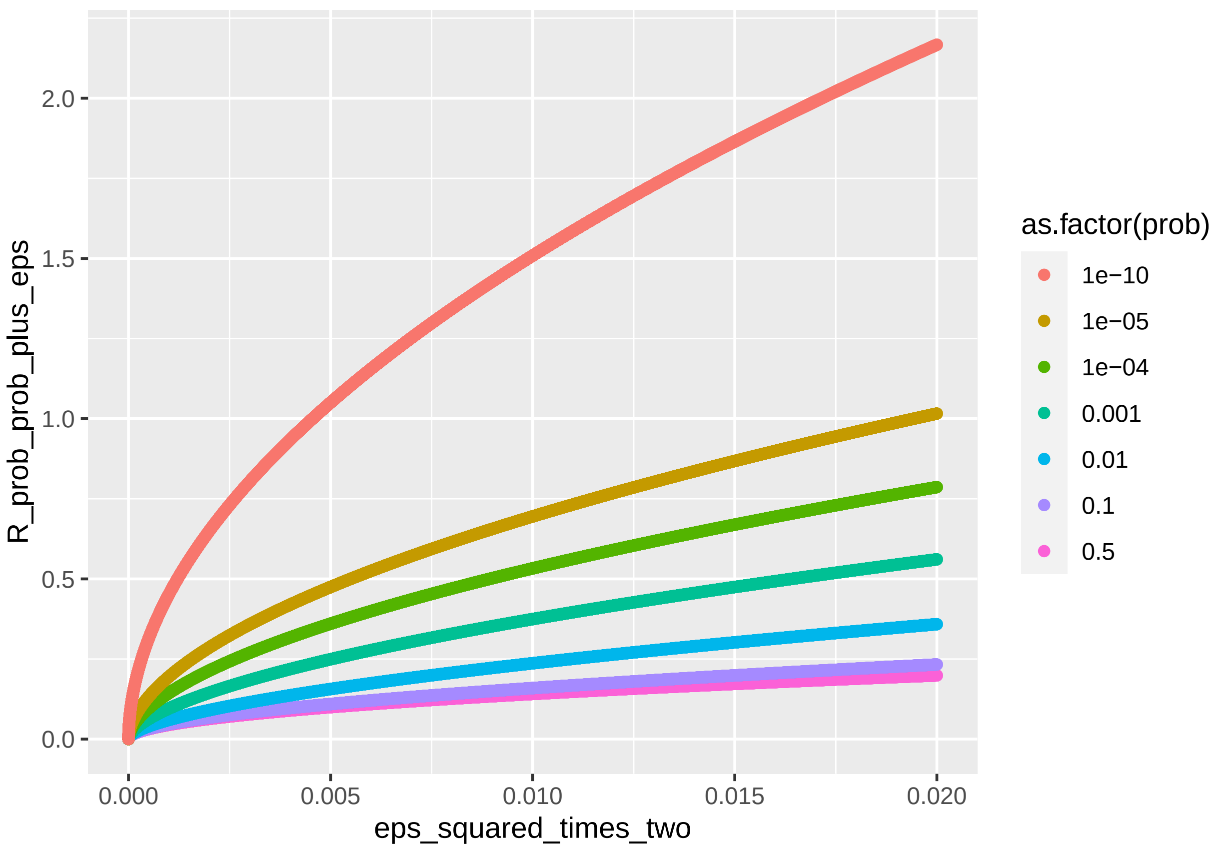
$$\begin{aligned}
g''(\epsilon) &= \frac{1}{p+\epsilon} + \frac{1}{1-p-\epsilon} - 4 \\
&\implies g''(\xi) \geq 0
\end{aligned}$$

where the last line follows because $\frac{1}{p} + \frac{1}{1-p} \geq 4$, for $p \in (0, 1)$.

Putting together the pieces, we see that we have the needed result.

Exercise 7

```r
library(magrittr)
epsilon <- (1:10000) / 100000
p <- c(0.5, 0.1, 0.01, 0.001, 0.0001, 0.00001, 1e-10)
tibble::tibble(eps = rep(epsilon, times = length(p)),
               prob = rep(p, each = length(epsilon))) %>%
  dplyr::mutate(eps_squared_times_two = 2 * eps ^ 2) %>%
  dplyr::mutate(R_prob_prob_plus_eps = (prob + eps) * log((prob + eps) / prob) -
                  (1 - prob - eps) * log((1 - prob - eps) / (1 - prob))
                ) %>%
  ggplot2::ggplot() +
  ggplot2::geom_point(mapping = ggplot2::aes(x = eps_squared_times_two,
                                             y = R_prob_prob_plus_eps, colour = as.factor(prob)))
```



Smaller values of p correspond to plots that, near $2\epsilon^2 = 0$, are steeper.

Exercise 8

We start with Hoeffding's bound.

$$
\begin{aligned}
0.01 = \mathbb{P}(|\frac{S_n}{n} - p| \geq \epsilon) &\leq 2e^{-2n\epsilon^2} \\
\implies \frac{1}{200} &\leq e^{-20000\epsilon^2} \\
\implies \frac{1}{20000}\log 200 &\geq \epsilon^2 \\
\implies \epsilon &= \sqrt{\frac{\log 200}{20000}} \approx 0.016
\end{aligned}
$$

Exercise 9

Taking majority vote and getting the wrong answer is equivalent to having $\hat{p}_n - p \geq \epsilon$. Note the absence of the absolute value specification.

We use Hoeffding, after adjusting for symmetry, to see that:

$$
\begin{aligned}
\mathbb{P}(\frac{S_n}{n} - p \geq \epsilon) &\leq e^{-2n\epsilon^2} \\
&= \delta \\
&\implies \delta \leq e^{-2n\epsilon^2} \\
&\implies \log \delta = -2n\epsilon^2 \\
&\implies n \geq \frac{\log \delta}{-2\epsilon^2}
\end{aligned}
$$