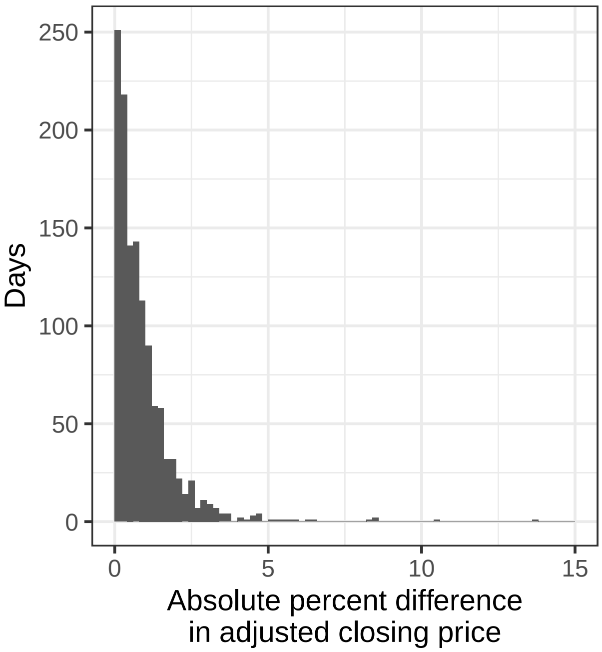
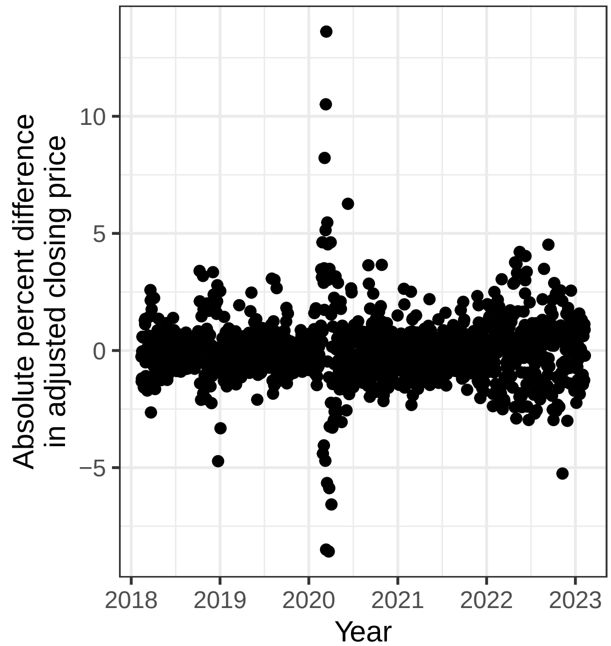
Announcements

- Make sure to sign in on the google form (I send a list of what section questions are useful for what pset questions afterwards)
- Pset 3 due Friday 2/17

Stonks

The following questions deal with the past 5 years of S&P 500 adjusted closing prices available here.





In this section, we will be modeling the day-to-day absolute percent differences in the adjusted closing price of the S&P 500 as $Y_1, ..., Y_n \sim \text{Expo}(\lambda)$.

1. Find the score of λ (recall that the score is $\frac{\partial}{\partial\lambda}\ell(\lambda;\vec{y})$) in terms of the sample mean and verify that $E(s(\lambda^*;\vec{Y})) = 0$.

2. Verify the information equality by showing $-E(s'(\lambda;\vec{Y})) = \text{Var}(s(\lambda;\vec{Y}))$.

3. Find the Fisher information $\mathcal{I}_{\vec{Y}}(\lambda^*)$. Then, find a function g such that $\mathcal{I}_{\vec{Y}}(g(\lambda^*))$ is constant (this is the variance stabilizing transformation of the Exponential distribution). Hint: Recall that the Fisher

information for a transformation is $\mathcal{I}_{\vec{Y}}(g(\lambda^*)) = \frac{\mathcal{I}_{\vec{Y}}(\lambda^*)}{g'(\lambda^*)^2}$.

4. Verify this is indeed the variance stabilizing transformation through simulation.

```
set.seed(111)

# TODO: Show the variance stabilization through simulation
```

5. Show that the MLE of $\hat{\lambda}$ is consistent for λ. That is, show that $\hat{\lambda} \to \lambda$ as $n \to \infty$ by showing the MSE goes to 0, a LLN holds, making a claim using the CMT, or showing convergence directly.

6. Find the asymptotic distribution of the MLE and its approximate distribution for large n.

7. In his book *The Black Swan*, Nassim Taleb argues that part of the reason for the 2008 financial crisis was a failure to model market fluctuations and assign sufficient probability to extreme events. Let us consider daily absolute differences above τ to be extreme events. Let $X_i = I(Y_i > \tau)$. Show that $\bar{X}$ is consistent for $p = P(Y_i > \tau)$ first by using MSE and then by using the law of large numbers.

8. Find the asymptotic distribution of $\bar{X}$ and its approximate distribution for large n in terms of λ.

9. Now, suppose we estimate $P(Y_i > \tau)$ with $\hat{p} = e^{-\hat{\lambda}\tau}$. Find the asymptotic distribution of $\hat{p}$ and its

approximate distribution for large n.

10. The distributions in 8 and 9 should have the same mean. However, the variances are different. Estimate the standard error of each estimator for the stocks data with $\tau = 5$. Explain why your results are what they are.

```
n <- length(stocks$close_diff)
tau = 5

# TODO: Find SE estimates for the estimators
```

11. Though $\hat{p}$ is more efficient, it might be less robust. Interpret the MSEs of the two estimators for estimating $P(Y > \tau)$ when $Y \sim \text{Expo}(0.5)$ and $Y \sim \text{Log-Normal}(0.1, 1)$. In this simulation, $\tau = 5$ and $n = 100$.

```
nsims <- 10^5
n <- 100
tau <- 5
mse_xbar_lnorm <- vector(length = nsims)
mse_phat_lnorm <- vector(length = nsims)
mse_xbar_exp <- vector(length = nsims)
mse_phat_exp <- vector(length = nsims)
for (i in 1:nsims) {
  log_norms <- rlnorm(n, 0.1, 1)
  mse_xbar_lnorm[i] <- (mean(log_norms > tau) - plnorm(tau, 0.1, 1, lower.tail = F))^2
  mse_phat_lnorm[i] <- (exp(-1/mean(log_norms) * tau) - plnorm(tau, 0.1, 1, lower.tail = F))^2

  expos <- rexp(n, 0.5)
  mse_xbar_exp[i] <- (mean(expos > tau) - pexp(tau, 0.5, lower.tail = F))^2
  mse_phat_exp[i] <- (exp(-1/mean(expos) * tau) - pexp(tau, 0.5, lower.tail = F))^2
}

output <- rbind(c(mean(mse_xbar_lnorm), mean(mse_phat_lnorm)),
      c(mean(mse_xbar_exp), mean(mse_phat_exp)))
colnames(output) <- c("Xbar", "phat")
rownames(output) <- c("Log Normal", "Expo")
round(output, digits = 5)
```

```
##                 Xbar    phat
## Log Normal 0.00061 0.00053
## Expo       0.00076 0.00041
```

Ty Mup

Ty Mup is taking an exam with n equally hard questions. He has a probability p_2 of getting each question right independently. However, there is also a $0 < p_1 < 1$ probability he sleeps through his alarm and misses the exam entirely. Let Y be the number of questions he gets right on his exam. (This distribution is called the zero-inflated binomial.)

1. Find $E(Y|Y > 0)$.

2. Unfortunately, the day is February 2nd in Punxsutawney and Ty is destined to repeat this day d times, scoring i.i.d Y_i on the exams. Find the likelihood function, the log-likelihood function, the score for p_1 and p_2. (Hint: Let m be the number of 0s.)

3. Find a two-dimensional sufficient statistic (a two dimensional statistic that contains all the information about the likelihood).

4. Find the Fisher information for p_1. (Hint: Write M as $M_1 + M_2$ where M_1 is the number of days Ty slept through the alarm and M_2 is the number of times he took the test and got a 0.)

5. Check that this Fisher information gives the correct result in the cases $p_2 = 1$ and $p_2 = 0$.

6. Let B be the event that Ty sleeps through the alarm at least once. Show that as $d \to \infty$,

$$I_B \frac{d^{1/2}(\bar{Y} - (1-p_1)np_2)}{\sqrt{np_2(1-p_2)(1-p_1) + (np_2)^2 p_1(1-p_1)}} \xrightarrow{d} \mathcal{N}(0,1)$$