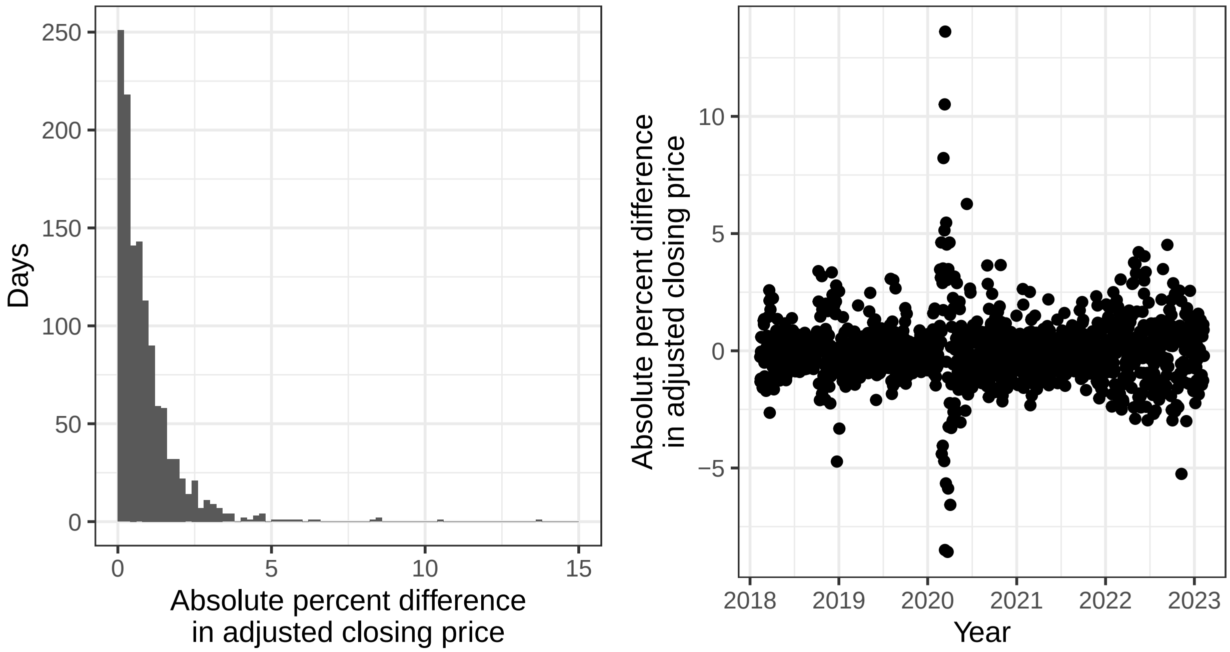
Announcements

- 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 $E(s(\lambda^*;\vec{Y})) = 0$.

The likelihood function is

$$L(\lambda;\vec{y}) = \prod_{i=1}^{n} \lambda e^{-\lambda y_i} = \lambda^n e^{-\lambda \sum_{i=1}^{n} y_i}$$

The log likelihood function is

$$\ell(\lambda;\vec{y}) = n\log(\lambda) - \lambda\sum_{i=1}^{n} y_i$$

The score is

$$s(\lambda;\vec{y}) = n/\lambda - n\bar{y} \implies E(s(\lambda^*;\vec{Y})) = n/\lambda^* - nE(\bar{Y}) = 0$$

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

$$E(s'(\lambda^*;\vec{Y})) = -n/\lambda^{*2}$$

$$\text{Var}(s(\lambda^*;\vec{Y})) = \text{Var}(n\bar{Y}) = n^2 \cdot \frac{1}{n\lambda^{*2}} = -E(s'(\lambda^*;\vec{Y}))$$

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

$$\mathcal{I}_{\vec{Y}}(\lambda^*) = \text{Var}(s(\lambda^*; \vec{Y})) = \frac{n}{\lambda^{*2}}$$

We need a function such that

$$\frac{n}{g'(\lambda^*)^2 \lambda^{*2}} \propto 1$$

This turns out to be about the easiest differential equation ever:

$$g'(\lambda^*) \propto \frac{1}{\lambda^*} \implies g(\lambda^*) \propto \log(\lambda^*)$$

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

```
set.seed(111)

sapply(c(0.001, 0.01, 0.1, 1, 10, 100, 1000), function(lambda) var(log(rexp(100000, lambda))))
```

```
## [1] 1.630103 1.651781 1.654339 1.629304 1.655330 1.641993 1.645989
```

Regardless of the rate of the exponential, taking the log gives the same variance.

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.

Setting the score to 0 gives $\hat{\lambda} = 1/\bar{y}$. We will show consistency with the continuous mapping theorem. Because $\bar{Y_n} \to 1/\lambda$ by the LLN, $\hat{\lambda} = 1/\bar{Y_n} \to \lambda$ by the CMT since $1/x$ is a continous function.

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

$$\sqrt{n}(\hat{\lambda} - \lambda^*) \to \mathcal{N}\left(0, \frac{1}{\mathcal{I}_1(\lambda^*)}\right)$$

which is $\mathcal{N}\left(0, \lambda^{*2}\right)$ by plugging in the Fisher information with $n = 1$. For large n, this gives the approximation $\hat{\lambda} \sim \mathcal{N}(\lambda^*, \lambda^{*2}/n)$.

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.

Since the Y_i are i.i.d., by the story of the binomial $n\bar{X} \sim \text{Bin}(n, p)$. Thus, $E(\bar{X}) = p$ and $\text{Var}(\bar{X}) = p(1-p)/n \to 0$ as $n \to \infty$. Since $\text{MSE}(\bar{X}, p) = \text{Bias}(\bar{X})^2 + \text{Var}(\bar{X}) \to 0$, $\bar{X}$ is a consistent estimator. This can be seen more easily from the fact that $\bar{X} \to E(X_i) = E(I(Y_i > \tau)) = p$ by the law of large numbers and fundamental bridge.

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

Using the Exponential CDF, $p = P(Y_i > \tau) = e^{-\lambda\tau}$. By the CLT,

$$\frac{\sqrt{n}(\bar{X} - e^{-\lambda\tau})}{\sqrt{e^{-\lambda\tau}(1 - e^{-\lambda\tau})}} \to \mathcal{N}(0, 1) \iff \bar{X} \sim \mathcal{N}(e^{-\lambda\tau}, \frac{e^{-\lambda\tau}(1 - e^{-\lambda\tau})}{n})$$

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.

We will use the Delta Method with $g(\lambda) = e^{-\lambda\tau}$. Differentiating gives $g'(\lambda) = -\tau e^{-\lambda\tau}$. Using the asymptotic distribution from 6 and the Delta Method,

$$\sqrt{n}(\hat{\lambda} - \lambda^*) \to \mathcal{N}\left(0, \lambda^{*2}\right) \implies \frac{\sqrt{n}(\hat{p} - e^{-\lambda^*\tau})}{|-\tau e^{-\lambda^*\tau}|} \to \mathcal{N}\left(0, \lambda^{*2}\right)$$

This gives the approximate distribution

$$\hat{p} \sim \mathcal{N}\left(e^{-\lambda^*\tau}, \frac{\tau^2 e^{-2\lambda^*\tau}\lambda^{*2}}{n}\right)$$

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.

```
lambda_hat <- 1/mean(abs(stocks$close_diff))
n <- length(stocks$close_diff)
tau = 5
round(c("Xbar SE" = sqrt(exp(-lambda_hat * tau) * (1-exp(-lambda_hat * tau)) / n),
  "phat SE" = sqrt(tau^2 * exp(-2 * lambda_hat * tau) * lambda_hat^2) / n), digits = 5)
```

```
## Xbar SE phat SE
## 0.00175 0.00002
```

We've made a distributional assumption about Y, so observing values even if they aren't over the threshold provides information, lowering the standard error of the estimate.

11. Though $\hat{p}$ is more efficient, it might be less robust. Find the MSEs of the two estimators for estimating $P(Y > \tau)$ when $Y \sim \text{Expo}(0.5)$ and $Y \sim \text{Log-Normal}(0.1, 1)$. Let $\tau = 5$ and $n = 100$, and perform 10^5 simulations.

```
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       0.00075 0.00041
```

It turns out that $\hat{p}$ has lower MSE on both the correct model and a misspecified model, so it is actually both more efficient and more robust. However, this is not always the case!

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 $E(Y|Y>0)$.

If $Y > 0$, he must have made it to the exam on time. Call this event A. From the problem description, $Y|A \sim \text{Bin}(n, p_2)$. Then, by the law of total expectation,

$$E(Y|A) = E(Y|A, Y>0)P(Y>0|A) + E(Y|A, Y=0)P(Y=0|A)$$

Since $Y > 0 \implies A$,

$$E(Y|Y>0) = E(Y|A, Y>0) = \frac{E(Y|A)}{P(Y>0|A)} = \frac{np_2}{1-(1-p_2)^n}$$

from the binomial PMF. This is slightly above np_2 as expected.

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.)

The likelihood can be written as the probability of observing the 0 values times the probability of observing everything else:

$$L(p_1, p_2; \vec{y}) = (p_1 + (1-p_1)(1-p_2)^n)^m \prod_{i=1, y_i \neq 0}^{d} (1-p_1)p_2^{y_i}(1-p_2)^{n-y_i}$$

Taking the log and using the fact that $\sum_{i=1, y_i\neq 0}^{d} y_i = d\bar{y}$ (where the mean is over all y_i, not just the positive ones),

$$\begin{aligned}
\ell(p_1, p_2; \vec{y}) &= m\log(p_1 + (1-p_1)(1-p_2)^n) + \sum_{i=1, y_i\neq 0}^{d} \log(1-p_1) + y_i\log(p_2) + (n-y_i)\log(1-p_2) \\
&= m\log(p_1 + (1-p_1)(1-p_2)^n) + (d-m)\log(1-p_1) + d\bar{y}\log(p_2) + \log(1-p_2)(n(d-m) - d\bar{y})
\end{aligned}$$

$$\begin{aligned}
s(p_1; \vec{y}) &= \frac{\partial}{\partial p_1}\ell(p_1, p_2; \vec{y}) = \frac{m(1-(1-p_2)^n)}{p_1 + (1-p_1)(1-p_2)^n} - \frac{d-m}{1-p_1} \\
s(p_2; \vec{y}) &= \frac{\partial}{\partial p_2}\ell(p_1, p_2; \vec{y}) = -\frac{nm(1-p_1)(1-p_2)^{n-1}}{p_1 + (1-p_1)(1-p_2)^n} + \frac{d\bar{y}}{p_2} - \frac{n(d-m) - d\bar{y}}{1-p_2}
\end{aligned}$$

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

$(m, \bar{y})$

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.)

$$\begin{aligned}
\text{Var}(M) &= E(\text{Var}(M|M_1)) + \text{Var}(E(M|M_1)) \\
&= E(\text{Var}(M_1 + M_2|M_1)) + \text{Var}(E(M_1 + M_2|M_1)) \\
&= E(\text{Var}(M_2|M_1)) + \text{Var}(M_1 + E(M_2|M_1))
\end{aligned}$$

since the variance of a constant is 0. Also, $M_2|M_1 \sim \text{Bin}(d - M_1, (1-p_2)^n)$ since there are $d - M_1$ days Ty actually took the test and a $(1-p_2)^n$ probability of getting a 0 if he took it. Using the mean and variance of a binomial,

$$\begin{aligned}\text{Var}(M) &= E((d-M_1)(1-p_2)^n(1-(1-p_2)^n)) + \text{Var}(M_1 + (d-M_1)(1-p_2)^n)\\ &= d(1-p_1)(1-p_2)^n(1-(1-p_2)^n) + (1-(1-p_2)^n)^2 dp_1(1-p_1)\end{aligned}$$

$$\begin{aligned}\mathcal{I}(p_1) &= \text{Var}(s(p_1;\vec{Y}))\\ &= \text{Var}\left(\frac{M(1-(1-p_2)^n)}{p_1+(1-p_1)(1-p_2)^n} - \frac{d-M}{1-p_1}\right)\\ &= \text{Var}\left(M\left(\frac{(1-(1-p_2)^n)}{p_1+(1-p_1)(1-p_2)^n} + \frac{1}{1-p_1}\right)\right)\\ &= \left(\frac{(1-(1-p_2)^n)}{p_1+(1-p_1)(1-p_2)^n} + \frac{1}{1-p_1}\right)^2 \text{Var}(M)\end{aligned}$$

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

For $p_2 = 1$,

$$\mathcal{I}(p_1) = \left(\frac{1}{p_1} + \frac{1}{1-p_1}\right)^2 dp_1(1-p_1) = \frac{d}{p_1(1-p_1)}$$

In this case, the only 0s come from missing the alarm, so the proportion of 0s $\hat{p_1}$ has the distribution $d\hat{p_1} \sim \text{Bin}(d, p_1)$. This means $\hat{p_1} \approx \mathcal{N}\left(p_1, \frac{p_1(1-p_1)}{d}\right)$, and the asymptotic approximate distribution of the MLE $\hat{p}$ using the Fisher information is equally $\mathcal{N}(p_1, 1/\mathcal{I}(p_1))$.

For $p_2 = 0$, $\mathcal{I}(p_1) = 0$ since $\text{Var}(M) = 0$. That is, our data set gives no information on p_1 since the Y_i were certain to be 0 regardless of what p_1 was. (Note that an asymptotic distribution for the MLE $\hat{p_2}$ can't be written in terms of the Fisher information here because the Fisher information is 0 and the regularity conditions specity it must be positive.)

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

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

Let A be the event Ty made it to the exam on time. By Law of Total Expectation

$$E(\bar{Y}) = E(Y) = E(Y|A)P(A) + E(Y|A)P(A^c) = np_2(1-p_1)$$

Likewise, by Eve's law,

$$\text{Var}(Y) = E(\text{Var}(Y|I_A)) + \text{Var}(E(Y|I_A)) = E(np_2(1-p_2)I_A) + \text{Var}(np_2 I_A) = np_2(1-p_2)(1-p_1) + (np_2)^2 p_1(1-p_1)$$

Therefore, $\text{Var}(\bar{Y}) = \text{Var}(Y)/d$.

By the CLT, $\frac{d^{1/2}(\bar{Y}-(1-p_1)dp_2)}{\sqrt{np_2(1-p_2)(1-p_1)+(np_2)^2 p_1(1-p_1)}} \to Z$ with $Z \sim \mathcal{N}(0,1)$. Also, $P(|I_B - 1| > \epsilon)$ is 0 if $\epsilon \geq 1$ and it is $(1-p_1)^d$ if $0 < \epsilon < 1$. Since $p_1 < 1$, $P(|I_B - 1| > \epsilon) \to 0$, so $I_B \xrightarrow{p} 1$. Thus, by Slutsky's theorem,

$$I_B \frac{d^{1/2}(\bar{Y} - (1-p_1)dp_2)}{\sqrt{np_2(1-p_2)(1-p_1) + (np_2)^2 p_1(1-p_1)}} \xrightarrow{d} 1 \cdot Z \sim \mathcal{N}(0,1)$$