STATS 200: Solutions to Homework 1

1. (a) Recall that a Binomial(n, p) random variable has mean np and variance $np(1-p)$. A total of pN people support Hillary, each voting independently with probability $\frac{1}{2}$, so $V_{\text{Hillary}} \sim \text{Binomial}(pN, \frac{1}{2})$. Then

$$\mathbb{E}[V_{\text{Hillary}}] = \tfrac{1}{2}pN, \;\; \text{Var}[V_{\text{Hillary}}] = \tfrac{1}{4}pN,$$

and the standard deviation of V_{Hillary} is $\sqrt{\frac{1}{4}pN}$. Similarly, as $(1-p)N$ people support Donald, $V_{\text{Donald}} \sim \text{Binomial}((1-p)N, \frac{1}{2})$, so

$$\mathbb{E}[V_{\text{Donald}}] = \tfrac{1}{2}(1-p)N, \;\; \text{Var}[V_{\text{Donald}}] = \tfrac{1}{4}(1-p)N,$$

and the standard deviation of V_{Donald} is $\sqrt{\frac{1}{4}(1-p)N}$. The fraction of voters who vote for Hillary is

$$\frac{V_{\text{Hillary}}}{V_{\text{Hillary}} + V_{\text{Donald}}} = \frac{V_{\text{Hillary}}/N}{V_{\text{Hillary}}/N + V_{\text{Donald}}/N}.$$

As $\mathbb{E}[V_{\text{Hillary}}/N] = \frac{1}{2}p$ (a constant) and $\text{Var}[V_{\text{Hillary}}/N] = \frac{1}{4}p/N \to 0$ as $N \to \infty$, V_{Hillary}/N should be close to $\frac{1}{2}p$ with high probability when N is large. Similarly, as $\mathbb{E}[V_{\text{Donald}}/N] = \frac{1}{2}(1-p)$ and $\text{Var}[V_{\text{Donald}}/N] = \frac{1}{4}(1-p)/N \to 0$ as $N \to \infty$, V_{Donald}/N should be close to $\frac{1}{2}(1-p)$ with high probability when N is large. Then the fraction of voters for Hillary should, with high probability, be close to

$$\frac{\frac{1}{2}p}{\frac{1}{2}p + \frac{1}{2}(1-p)} = p.$$

(The above statements "close to with high probability" may be formalized using Chebyshev's inequality, which states that a random variable is, with high probability, not too many standard deviations away from its mean.)

(b) Let $V_{\text{H,p}}$ and $V_{\text{H,a}}$ be the number of passive and active voters who vote for Hillary, and similarly define $V_{\text{D,p}}$ and $V_{\text{D,a}}$ for Donald. There are $q_H pN$ passive Hillary supporters, each of whom vote independently with probability $\frac{1}{4}$, so

$$V_{\text{H,p}} \sim \text{Binomial}(q_H pN, \tfrac{1}{4}).$$

Similarly,

$$\begin{aligned} V_{\text{H,a}} &\sim \text{Binomial}((1-q_H)pN, \tfrac{3}{4}), \\ V_{\text{D,p}} &\sim \text{Binomial}(q_D(1-p)N, \tfrac{1}{4}), \\ V_{\text{D,a}} &\sim \text{Binomial}((1-q_D)(1-p)N, \tfrac{3}{4}), \end{aligned}$$

and these four random variables are independent. Since $V_{\text{Hillary}} = V_{\text{H,p}} + V_{\text{H,a}}$,

$$\begin{aligned} \mathbb{E}[V_{\text{Hillary}}] &= \mathbb{E}[V_{\text{H,p}}] + \mathbb{E}[V_{\text{H,a}}] = \tfrac{1}{4}q_H pN + \tfrac{3}{4}(1-q_H)pN, \\ \text{Var}[V_{\text{Hillary}}] &= \text{Var}[V_{\text{H,p}}] + \text{Var}[V_{\text{H,a}}] = \tfrac{3}{16}q_H pN + \tfrac{3}{16}(1-q_H)pN = \tfrac{3}{16}pN, \end{aligned}$$

and the standard deviation of V_{Hillary} is $\sqrt{\frac{3}{16}pN}$. Similarly,

$$\begin{aligned} \mathbb{E}[V_{\text{Donald}}] &= \mathbb{E}[V_{\text{D,p}}] + \mathbb{E}[V_{\text{D,a}}] = \tfrac{1}{4}q_D(1-p)N + \tfrac{3}{4}(1-q_D)(1-p)N, \\ \text{Var}[V_{\text{Donald}}] &= \text{Var}[V_{\text{D,p}}] + \text{Var}[V_{\text{D,a}}] = \tfrac{3}{16}q_D(1-p)N + \tfrac{3}{16}(1-q_D)(1-p)N \\ &= \tfrac{3}{16}(1-p)N, \end{aligned}$$

and the standard deviation of V_{Donald} is $\sqrt{\frac{3}{16}(1-p)N}$.

The quantity $\hat{p}$ estimates p, but in this case p may not be the fraction of voters who vote for Hillary: By the same argument as in part (a), the fraction of voters who vote for Hillary is given by

$$\begin{aligned} \frac{V_{\text{Hillary}}}{V_{\text{Hillary}} + V_{\text{Donald}}} &= \frac{V_{\text{Hillary}}/N}{V_{\text{Hillary}}/N + V_{\text{Donald}}/N} \\ &\approx \frac{\frac{1}{4}q_H p + \frac{3}{4}(1-q_H)p}{\frac{1}{4}q_H p + \frac{3}{4}(1-q_H)p + \frac{1}{4}q_D(1-p) + \frac{3}{4}(1-q_D)(1-p)}, \end{aligned}$$

where the approximation is accurate with high probability when N is large. When $q_H \neq q_D$, this is different from p: For example, if $q_H = 0$ and $q_D = 1$, this is equal to $\frac{p}{p+(1-p)/3}$ which is greater than p, reflecting the fact that Hillary supporters are more likely to vote than are Donald supporters.

(c) Let $\hat{p}$ be the proportion of the 1000 surveyed people who support Hillary. Among the surveyed people supporting Hillary, let $\hat{q}_H$ be the proportion who are passive. Similarly, among the surveyed people supporting Donald, let $\hat{q}_D$ be the proportion who are passive. (Note that these are observed quantities, computed from our sample of 1000 people.) Then we may estimate the number of voters for Hillary and Donald by

$$\begin{aligned} \hat{V}_{\text{Hillary}} &= \tfrac{1}{4}\hat{q}_H\hat{p}N + \tfrac{3}{4}(1-\hat{q}_H)\hat{p}N \\ \hat{V}_{\text{Donald}} &= \tfrac{1}{4}\hat{q}_D(1-\hat{p})N + \tfrac{3}{4}(1-\hat{q}_D)(1-\hat{p})N. \end{aligned}$$

$\hat{q}_H\hat{p}$ is simply the proportion of the 1000 surveyed people who both support Hillary and are passive. Hence, letting $X_1, \ldots, X_{1000}$ indicate whether each surveyed person both supports Hillary and is passive, we have

$$\hat{q}_H\hat{p} = \tfrac{1}{n}(X_1 + \ldots + X_n).$$

Each $X_i \sim \text{Bernoulli}(q_H p)$, so linearity of expectation implies $\mathbb{E}[\hat{q}_H\hat{p}] = q_H p$. Similarly, $(1-\hat{q}_H)\hat{p}$, $\hat{q}_D(1-\hat{p})$, and $(1-\hat{q}_D)(1-\hat{p})$ are the proportions of the 1000 surveyed people who support Hillary and are active, support Donald and are passive, and support Donald and are active, so the same argument shows $\mathbb{E}[(1-\hat{q}_H)\hat{p}] = (1-q_H)p$, $\mathbb{E}[\hat{q}_D(1-\hat{p})] = q_D(1-p)$, and $\mathbb{E}[(1-\hat{q}_D)(1-\hat{p})] = (1-q_D)(1-p)$. Then applying linearity of expectation again yields

$$\mathbb{E}[\hat{V}_{\text{Hillary}}] = \mathbb{E}[V_{\text{Hillary}}], \quad \mathbb{E}[\hat{V}_{\text{Donald}}] = \mathbb{E}[V_{\text{Donald}}].$$

2. X has the same distribution as $-X$, so $\mathbb{E}[X] = \mathbb{E}[-X] = -\mathbb{E}[X]$, hence $\mathbb{E}[X] = 0$. Similarly $\mathbb{E}[Y] = 0$. Also (X, Y) has the same joint distribution as $(-X, Y)$, so $\mathbb{E}[XY] = \mathbb{E}[-XY] = -\mathbb{E}[XY]$, hence $\mathbb{E}[XY] = 0$. Then $\text{Cov}[X, Y] = \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y] = 0$. On the other hand, conditional on $X = x$, Y is uniformly distributed on the interval $[-\sqrt{1-x^2}, \sqrt{1-x^2}]$. As this depends on x, X and Y are not independent.

3. X is the sum of n independent Bernoulli random variables $X_1, \ldots, X_n$, each with moment generating function

$$M_{X_i}(t) = \mathbb{E}\exp(tX_i) = pe^t + (1-p).$$

Combining these and applying independence yields

$$M_X(t) = \mathbb{E}\exp(tX) = \mathbb{E}\exp(t(X_1 + \ldots + X_n)) = \prod_{i=1}^{n} \mathbb{E}\exp(tX_i) = (1 - p + pe^t)^n.$$

4. (a) Any linear combination of $Y_1, \ldots, Y_m$ is a linear combination of $X_1, \ldots, X_k$, so $(Y_1, \ldots, Y_m)$ is multivariate normal. Using linearity of expectation and bilinearity of covariance, we compute

$$\begin{aligned}
\mathbb{E}[Y_i] &= a_{i1}\,\mathbb{E}[X_1] + \ldots + a_{ik}\,\mathbb{E}[X_k] = 0, \\
\text{Cov}[Y_i, Y_j] &= \text{Cov}[a_{i1}X_1 + \ldots + a_{ik}X_k,\ a_{j1}X_1 + \ldots + a_{jk}X_k] \\
&= \sum_{r=1}^{k}\sum_{s=1}^{k} a_{ir}a_{js}\,\text{Cov}[X_r, X_s] = \sum_{r=1}^{k}\sum_{s=1}^{k} a_{ir}a_{js}\Sigma_{rs} = a_i\Sigma a_j^T,
\end{aligned}$$

where a_i and a_j denote rows i and j of the matrix A. (The computation of covariance is valid for both $i \neq j$ and $i = j$; in the latter case this yields $\text{Var}[Y_i] = \text{Cov}[Y_i, Y_i]$.) As $a_i\Sigma a_j^T = (A\Sigma A^T)_{ij}$, this implies by definition $Y \sim \mathcal{N}(0, A\Sigma A^T)$.

(b) Take $X_1, \ldots, X_k \overset{iid}{\sim} \mathcal{N}(0,1)$, define $Z_j = a_{j1}X_1 + \ldots + a_{jk}X_k$ for each $j = 1, \ldots, k$, and let $Y_j = Z_j + \mu_j$. As $(X_1, \ldots, X_k)$ have the multivariate normal distribution $\mathcal{N}(0, I)$ where I is the $k \times k$ identity matrix, and as $AIA^T = AA^T = \Sigma$, part (a) implies $(Z_1, \ldots, Z_k) \sim \mathcal{N}(0, \Sigma)$. Then $(Y_1, \ldots, Y_k) \sim \mathcal{N}(\mu, \Sigma)$ (since adding the vector $\mu = (\mu_1, \ldots, \mu_k)$ does not change the variances and covariances of $Y_1, \ldots, Y_k$ but shifts their means by $\mu_1, \ldots, \mu_k$).

5. We have

$$\mathbb{P}[X + Y > 0 \mid X > 0] = \frac{\mathbb{P}[X + Y > 0, X > 0]}{\mathbb{P}[X > 0]}.$$

Since X has the same distribution as $-X$, $\mathbb{P}[X > 0] = \mathbb{P}[-X > 0] = \mathbb{P}[X < 0]$, so $\mathbb{P}[X > 0] = \frac{1}{2}$. To compute $\mathbb{P}[X + Y > 0, X > 0]$, note that this is the integral of the bivariate normal PDF $f_{X,Y}(x, y)$ in the region to the right of the y-axis and above the line $y = -x$. The integral of $f_{X,Y}(x, y)$ over all of $\mathbb{R}^2$ must equal 1; hence by rotational symmetry of $f_{X,Y}(x, y)$ around the origin, the integral over any wedge formed by two rays extending from the origin is $\theta/(2\pi)$ where θ is the angle formed by these rays. For the above region, this angle is $3\pi/4$, so $\mathbb{P}[X + Y > 0, X > 0] = 3/8$. Then $\mathbb{P}[X + Y > 0 \mid X > 0] = 3/4$.

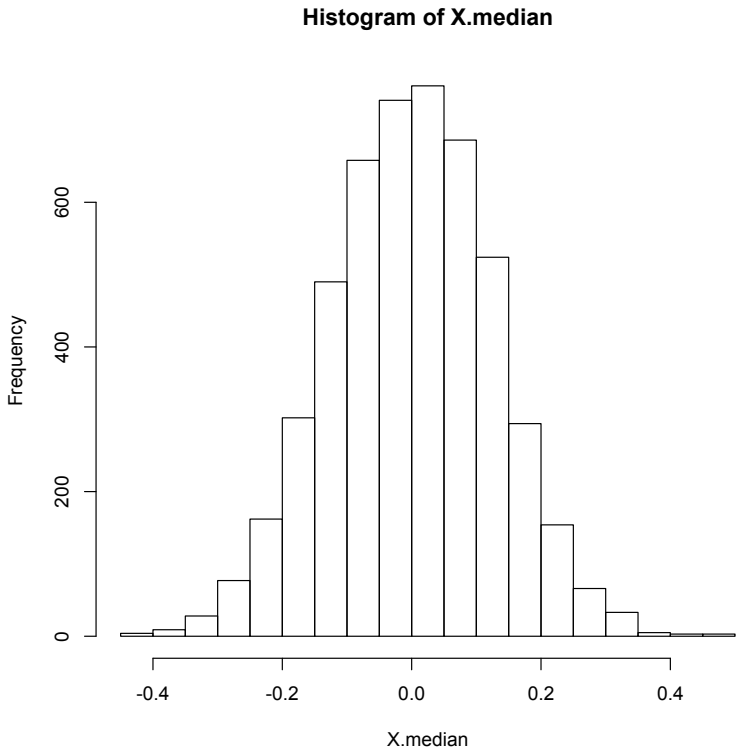
6. Code is as follows:

```
X.median = numeric(5000)

for(i in 1:5000) {
  X = rnorm(99, mean = 0, sd = 1)
  X.median[i] = median(X)
}

print(mean(X.median))
print(sd(X.median))
hist(X.median)
```

The mean and standard deviation of the sample medians are -0.001 and 0.126, respectively. The histogram of the medians is shown in the following figure. The sampling distribution of the sample median looks approximately normally distributed with this mean and standard deviation.


Histogram of X.median
Frequency
0
200
400
600
-0.4
-0.2
0.0
0.2
0.4
X.median