MATH 563 HOMEWORK 1
SOLUTIONS

SONJA PETROVIĆ

GRADING SCALE: 5 points per problem. About half the points given for correct mathematical result, and the other half for reasoning / explanation of the solution.

1. (Problem 5.1.)

Let $X = \#$ color blind people in a sample of size n. Then $X \sim \text{binomial}(n, p)$, where $p = .01$. The probability that a sample contains a color blind person is $P(X > 0) = 1 - P(X = 0)$, where $P(X = 0) = \binom{n}{0}(.01)^0(.99)^n = .99^n$. Thus,

$$P(X > 0) = 1 - .99^n > .95 \Leftrightarrow n > \log(.05)/\log(.99) \approx 299.$$

2. (Exercise 5.15.)

a.

$$\bar{X}_{n+1} = \frac{\sum_{i=1}^{n+1} X_i}{n+1} = \frac{X_{n+1} + \sum_{i=1}^{n} X_i}{n+1} = \frac{X_{n+1} + n\bar{X}_n}{n+1}.$$

b.

$$\begin{aligned}
nS_{n+1}^2 &= \frac{n}{(n+1)-1}\sum_{i=1}^{n+1}\left(X_i - \bar{X}_{n+1}\right)^2 \\
&= \sum_{i=1}^{n+1}\left(X_i - \frac{X_{n+1}+n\bar{X}_n}{n+1}\right)^2 && \text{(use (a))} \\
&= \sum_{i=1}^{n+1}\left(X_i - \frac{X_{n+1}}{n+1} - \frac{n\bar{X}_n}{n+1}\right)^2 \\
&= \sum_{i=1}^{n+1}\left[\left(X_i - \bar{X}_n\right) - \left(\frac{X_{n+1}}{n+1} - \frac{\bar{X}_n}{n+1}\right)\right]^2 && (\pm \bar{X}_n) \\
&= \sum_{i=1}^{n+1}\left[\left(X_i - \bar{X}_n\right)^2 - 2\left(X_i - \bar{X}_n\right)\left(\frac{X_{n+1}-\bar{X}_n}{n+1}\right) + \frac{1}{(n+1)^2}\left(X_{n+1}-\bar{X}_n\right)^2\right] \\
&= \sum_{i=1}^{n}\left(X_i - \bar{X}_n\right)^2 + \left(X_{n+1} - \bar{X}_n\right)^2 - 2\frac{\left(X_{n+1}-\bar{X}_n\right)^2}{n+1} + \frac{n+1}{(n+1)^2}\left(X_{n+1} - \bar{X}_n\right)^2 \\
& && \left(\text{since } \sum_1^n (X_i - \bar{X}_n) = 0\right) \\
&= (n-1)S_n^2 + \frac{n}{n+1}\left(X_{n+1} - \bar{X}_n\right)^2.
\end{aligned}$$

Date: Spring 2021.

3. (Exercise 5.4.) Let $X_i|P \sim_{iid} Bernoulli(P)$, for $i = 1, \ldots, n$, and let $P \sim Uniform(0,1)$.
   a) By the definition of Bernoulli distribution, $\Pr(X_i = x_i|P = p) = p^{x_i}(1-p)^{1-x_i}$. Thus by the independence of X_is we have

$$\Pr(X_1 = x_1, X_2 = x_2, \cdots, X_k = x_k|P = p)$$
$$= \prod_{i=1}^{k} p^{x_i}(1-p)^{1-x_i} = p^{\sum_{i=1}^k x_i} \cdot (1-p)^{\sum_{i=1}^k (1-x_i)} = p^t \cdot (1-p)^{(k-t)}$$

   Here $t = \sum_{i=1}^k x_i$.
   We know that for probability density functions, $f(x) = \int_{-\infty}^{\infty} f(x,y)dy = \int_{-\infty}^{\infty} f(x|y)f(y)dy$, also $P \sim U(0,1)$, thus we have

$$\Pr(X_1 = x_1, X_2 = x_2, \cdots, X_k = x_k)$$
$$= \int_{-\infty}^{\infty} \Pr(X_1 = x_1, X_2 = x_2, \cdots, X_k = x_k|P = p) f_P(p)dp$$
$$= \int_0^1 \Pr(X_1 = x_1, X_2 = x_2, \cdots, X_k = x_k|P = p) \cdot 1dp$$
$$= \int_0^1 p^t \cdot (1-p)^{(k-t)}dp = \frac{\Gamma(t+1)\Gamma(k-t+1)}{\Gamma(k+2)} = \frac{t!(k-t)!}{(k+1)!}$$

   b) Because usually $\int_a^b f(x)dx \cdot \int_a^b g(x)dx \neq \int_a^b f(x)g(x)dx$, thus
   $\prod_{i=1}^n \Pr(X_i = x_i) = \prod_{i=1}^n \int_0^1 p^{x_i}(1-p)^{(1-x_i)}dp$
   $\neq \int_0^1 \prod_{i=1}^n p^{x_i}(1-p)^{(1-x_i)}dp = \Pr(X_1 = x_1, \cdots, X_n = x_n)$.
4. (either Exercise 5.13 or Exercise 5.16)

5.13

$$\mathrm{E}\left(c\sqrt{S^2}\right) = c\sqrt{\frac{\sigma^2}{n-1}}\mathrm{E}\left(\sqrt{\frac{S^2(n-1)}{\sigma^2}}\right)$$
$$= c\sqrt{\frac{\sigma^2}{n-1}}\int_0^\infty \sqrt{q}\frac{1}{\Gamma\left(\frac{n-1}{2}\right)2^{(n-1)/2}}q^{\left(\frac{n-1}{2}\right)-1}e^{-q/2}dq,$$

Since $\sqrt{S^2(n-1)/\sigma^2}$ is the square root of a χ^2 random variable. Now adjust the integrand to be another χ^2 pdf and get

$$\mathrm{E}\left(c\sqrt{S^2}\right) = c\sqrt{\frac{\sigma^2}{n-1}} \cdot \frac{\Gamma(n/2)2^{n/2}}{\Gamma((n-1)/2)2^{((n-1)/2}} \underbrace{\int_0^\infty \frac{1}{\Gamma(n/2)2^{n/2}}q^{(n-1)/2} - \frac{1}{2}e^{-q/2}dq}_{=1 \text{ since } \chi_n^2 \text{ pdf}}.$$

So $c = \frac{\Gamma\left(\frac{n-1}{2}\right)\sqrt{n-1}}{\sqrt{2}\Gamma\left(\frac{n}{2}\right)}$ gives $\mathrm{E}(cS) = \sigma$.

5.16

Since $X_i \sim N(i, i^2), i = 1,2,3$ are independent, we denote $Z_i = \frac{X_i - i}{i}$, then $Z_i \sim N(0,1)$, i.i.d. We will use Z_1, Z_2, Z_3 to construct the required random variables.

(a) From Lemma 5.3.2.a, $Z_i^2 \sim \chi^2(1), i = 1,2,3$, then from Lemma 5.3.2.b, $Z_1^2 + Z_2^2 + Z_3^2 \sim \chi^2(1+1+1) = \chi^2(3)$, which is the desired random variable.

(b) There are two ways to construct this random variable. Firstly, a $t(2)$ distributed variable can be constructed as the ratio $\frac{U}{\sqrt{V/2}}$, where $U \sim N(0,1)$ and $V \sim \chi^2(2)$. We take $U = Z_1, V = Z_2^2 + Z_3^2$ to obtain $\frac{Z_1}{\sqrt{(Z_2^2+Z_3^2)/2}} \sim t(2)$.

Alternatively, one can also directly use definition 5.3.4 to construct $\frac{\bar{Z}}{S/\sqrt{3}} \sim t(3-1) = t(2)$, where $\bar{Z} = \frac{1}{3}(Z_1 + Z_2 + Z_3)$ and $S = \sqrt{\frac{Z_1^2+Z_2^2+Z_3^2-3\bar{Z}}{2}}$.

Remark: Please note that $S^2 \neq \frac{1}{3}(Z_1^2 + Z_2^2 + Z_3^2)$. There are some confusions in your homeworks.

(c) Since $Z_1^2 \sim \chi^2(1), Z_2^2 + Z_3^2 \sim \chi^2(2)$, we have $\frac{Z_1^2}{(Z_2^2+Z_3^2)/2} \sim F(1,2)$.