

Chapter 5: Properties of Random Samples

November 22, 2021

Exercise 5.1

The samples are drawn from Bernoulli trial with success rate 0.01. The probability of n samples not containing color-blind is 0.99^n . We want to find N such that for $n \geq N$, $0.99^n \leq 1 - 0.95$. $N \approx 299$. ■

Exercise 5.2

(a) Let T be the number of years until the first year's rainfall is exceeded. Then

$$\begin{aligned} P(T = k) &= P(X_2 \leq X_1, \dots, X_{k-1} \leq X_1, X_k > X_1) \\ &= \int_x P(X_2 \leq x, \dots, X_{k-1} \leq x, X_k > x | X_1 = x) f(x) dx \\ &= \int_x P(X_k > x) f(x) \prod_{i=2}^{k-1} P(X_i \leq x) dx \\ &= \int_x (1 - F(x)) f(x) F(x)^{k-1} dx \\ &= \int_x F(x)^{k-1} f(x) dx - \int_x F(x)^k f(x) dx \\ &= \frac{1}{k} F(x)^k \Big|_{-\infty}^{\infty} - \frac{1}{k+1} F(x)^{k+1} \Big|_{-\infty}^{\infty} \\ &= \frac{1}{k} - \frac{1}{k+1} \\ &= \frac{1}{k(k+1)} \end{aligned}$$
 ■

(b)

$$ET = \sum_k kP(T = k) = \sum_k \frac{1}{k+1} = \infty$$

■

Exercise 5.3

Since $\{X_i\}$ are i.i.d $\sim F_X(x)$ and Y_i is hierarchical wrt to X_i . $Y_i \sim \text{Bernoulli}(P(X_i > \mu)|X_i)$. So Y_i are i.i.d. Therefore the sum of $Y_i \sim \text{Binomial}(n, P(X_i > \mu)) = \text{Binomial}(n, 1 - F_X(\mu))$.

$$P(Y_i = k) = \binom{n}{k} (1 - F_X(\mu))^k F_X(\mu)^{n-k}$$

■

Exercise 5.4

(a) $X_i|P \sim \text{Bernoulli}(P)$ are i.i.d and $P \sim \text{Uniform}(0, 1)$. Let $T = \sum_{i=1}^k X_i$.

$$\begin{aligned} P(X_1 = x_1, \dots, X_k = x_k) &= \int_0^1 P(X_1 = x_1, \dots, X_k = x_k | P = p) f(p) dp \\ &= \int_0^1 \prod_{i=1}^k P(X_i = x_i | P = p) f(p) dp, \text{ Since } X_i|P \text{ are i.i.d} \\ &= \int_0^1 \prod_{i=1}^k p^{x_i} (1-p)^{1-x_i} f(p) dp \\ &= \int_0^1 p^{\sum_i x_i} (1-p)^{1-\sum_i x_i} f(p) dp \\ &= \int_0^1 p^t (1-p)^{1-t} f(p) dp \end{aligned}$$

where $t = \sum_i x_i$.

■

(b) From (a),

$$P(X_1 = x_1, \dots, X_n = x_n) = \int_0^1 p^t (1-p)^{1-t} f(p) dp$$

where $t = \sum_i^n x_i$.

On the other hand,

$$\prod_i^n P(X_i = x_i) = \prod_i^n \int_0^1 P(X_i = x_i | P = p) f(p) dp = \prod_i^n \int_0^1 p^{x_i} (1-p)^{1-x_i} dp$$

Therefore $P(X_1 = x_1, \dots, X_n = x_n) \neq \prod_i^n P(X_i = x_i)$. ■

Exercise 5.5

let $Y = \sum_i X_i$ then $\bar{X} = Y/n$. Suppose we have $f_Y(y)$, then

$$f_{\bar{X}}(\bar{x}) = f_Y(y) = f_Y(n\bar{x}) \left| \frac{dy}{d\bar{x}} \right| = n f_Y(n\bar{x})$$

Exercise 5.6

*Book has typos, it should be 5.2.9 instead of 5.2.3.

(a) Let $Z = X + Y$, $V = X$, then

$$f_{V,Z}(v, z) = f_{X,Y}(v, z - v) \left| \frac{\partial(X, Y)}{\partial(V, Z)} \right| = f_{X,Y}(v, z - v) \begin{vmatrix} 1 & 0 \\ -1 & 1 \end{vmatrix} = f_{X,Y}(v, z - v)$$

Therefore

$$f_Z(z) = \int_v f_{X,Y}(v, z - v) dv = \int_v f_X(v) f_Y(z - v) dv$$

(b) Let $Z = XY$, $V = X$, then

$$f_{V,Z}(v, z) = f_{X,Y}(v, z/v) \left| \frac{\partial(X, Y)}{\partial(V, Z)} \right| = f_{X,Y}(v, z/v) \begin{vmatrix} 1 & 0 \\ -z/v^2 & 1/v \end{vmatrix} = f_{X,Y}(v, z - v) \left| \frac{1}{v} \right|$$

Therefore

$$f_Z(z) = \int_v f_{X,Y}(v, z/v) dv = \int_v f_X(v) f_Y(z/v) \left| \frac{1}{v} \right| dv$$

(c) Let $Z = X/Y$, $V = X$, then

$$\left| \frac{\partial(X, Y)}{\partial(V, Z)} \right| = \begin{vmatrix} 1 & 0 \\ 1/z & -v/z^2 \end{vmatrix} = \left| \frac{v}{z^2} \right|$$

Therefore

$$f_Z(z) = \int_v f_{V,Z}(v, z) dv = \int_v f_{X,Y}(v, v/z) \left| \frac{\partial(X, Y)}{\partial(V, Z)} \right| dv = \int_v f_X(v) f_Y(v/z) \left| \frac{v}{z^2} \right| dv$$

Exercise 5.7

(a) Combining the terms on the right side and order the term by power of w , we get

$$\begin{aligned}\left(\frac{A}{\tau^2} - \frac{C}{\sigma^2}\right)w^3 &= 0 \\ \left(-\frac{2Az}{\tau^2} + \frac{B}{\tau^2} - \frac{D}{\sigma^2}\right)w^2 &= 0 \\ \left(A + \frac{Az^2}{\tau^2} - \frac{2Bz}{\tau^2} - C\right)w &= 0 \\ B + \frac{Bz^2}{\tau^2} - D &= 1\end{aligned}$$

We get linear equation of

$$\begin{pmatrix} \sigma^2 & 0 & -\tau^2 & 0 \\ -2z\sigma^2 & \sigma^2 & 0 & -\tau^2 \\ \tau^2 + z^2 & -2z & -\tau^2 & 0 \\ 0 & \tau^2 + z^2 & 0 & -\tau^2 \end{pmatrix} \begin{pmatrix} A \\ B \\ C \\ D \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \tau^2 \end{pmatrix}$$

The determinant is $(-\sigma^2 + \tau^2 + z^2)^2 + 4z^2\sigma^2 \neq 0$. So A, B, C, D exists.

(b) Skipping the trivial calculation.

Exercise 5.8

(a)

$$\begin{aligned}(n-1)S^2 &= \sum_i (X_i - \bar{X})^2 \\ &= \sum_i \left(X_i - \frac{1}{n} \sum_j X_j\right)^2 \\ &= \sum_i \left(X_i - \frac{2}{n} X_i \sum_j X_j + \frac{1}{n^2} \left(\sum_j X_j\right)^2\right) \\ &= \sum_i X_i - \frac{2}{n} \sum_i X_i \sum_j X_j + \frac{1}{n^2} \sum_i \left(\sum_j X_j\right)^2 \\ &= \sum_i X_i - \frac{2}{n} \sum_i \sum_j X_i X_j + \frac{1}{n} \sum_i \sum_j X_i X_j\end{aligned}$$

Multiply both side by $2n$, we get

$$\begin{aligned}
2n(n-1)S^2 &= 2n \sum_i X_i - 2 \sum_i \sum_j X_i X_j \\
&= n \sum_i X_i - 2 \sum_i \sum_j X_i X_j + n \sum_i X_i \\
&= n \sum_i X_i - 2 \sum_i \sum_j X_i X_j + n \sum_j X_j \\
&= \sum_j \sum_i X_i - 2 \sum_i \sum_j X_i X_j + \sum_i \sum_j X_j, \text{ (Note that } n = \sum_i 1 = \sum_j 1 \text{)} \\
&= \sum_i \sum_j (X_i - X_j)^2
\end{aligned}$$

■

(b) Let $Y_i = X_i - \theta_1$. Then $\mathbf{E}Y_i = 0$, $\mathbf{E}Y_i^j = \theta_j$.

$$\begin{aligned}
2N(N-1)S^2 &= \sum_i \sum_j (X_i - X_j)^2 \\
&= \sum_i \sum_j (Y_i - Y_j)^2 \\
&= \sum_{i \neq j} (Y_i)^2 - Y_i Y_j + (Y_j)^2 \\
2N(N-1)\mathbf{E}S^2 &= \sum_{i \neq j} E(Y_i)^2 - 2\mathbf{E}Y_i \mathbf{E}Y_j + \mathbf{E}(Y_j)^2 \\
&= [E(Y_1)^2 - 2\mathbf{E}Y_1 \mathbf{E}Y_2 + E(Y_2)^2] \\
&= 2N(N-1)\theta_2 \\
\mathbf{E}S^2 &= \theta_2
\end{aligned}$$

For $i \neq j$, i has N choices and j has $N-1$. Therefore there are $N(N-1)$ terms in the sum.

$$\begin{aligned}
(4N^2(N-1)^2)S^4 &= \sum_i \sum_j (Y_i - Y_j)^2 \sum_m \sum_n (Y_m - Y_n)^2 \\
&= \sum_{i \neq j} \sum_{m \neq n} (Y_i^2 - 2Y_i Y_j + Y_j^2)(Y_m^2 - 2Y_m Y_n + Y_n^2) \\
&= \sum_{i \neq j, m \neq n} Y_i^2 Y_m^2 - 2Y_i^2 Y_m Y_n + Y_i^2 Y_n^2 \\
&\quad - 2Y_i Y_j Y_m^2 + 4Y_i Y_j Y_m Y_n - 2Y_i Y_j Y_n^2 \\
&\quad + Y_j^2 Y_m^2 - 2Y_j^2 Y_m Y_n + Y_j^2 Y_n^2
\end{aligned}$$

We have terms of 3 patterns $Y_i^2 Y_m Y_n$, $Y_i^2 Y_m^2$ and $Y_i Y_j Y_m Y_n$. The rest are equivalent. Grouping them together, we have.

$Y_i^2 Y_m^2$ can be split into Y_i^4 when $i = m$ and $Y_i^2 Y_m^2$ when $m \neq i$

$Y_i Y_j Y_m Y_n$ will not vanish only when $i = m, j = n$ or $i = n, j = m$. It can be written as $2Y_i^2 Y_m^2$ when $i = m$ (Times 2 to account for both cases)

$Y_i^2 Y_m Y_n$ will vanish when we take the expected value since $m \neq n$ and $\mathbf{E}Y_i = 0$

$$\begin{aligned} \mathbf{E}(4N^2(N-1)^2)S^4 &= \mathbf{E} \sum_{i \neq j, m \neq n} 4Y_i^2 Y_m^2 - 8Y_i^2 Y_m Y_n + 4Y_i Y_j Y_m Y_n \\ &= \mathbf{E} \left[\sum_{i \neq j, m \neq n} 4Y_i^2 Y_m^2 + 4Y_i Y_j Y_m Y_n \right] \\ &= \mathbf{E} \left[4 \sum_{i \neq j, m \neq n, i=m} Y_i^4 + 4 \sum_{i \neq j, m \neq n, i \neq m} Y_i^2 Y_m^2 + 8 \sum_{i \neq j, m \neq n, i=m, j=n} Y_i^2 Y_m^2 \right] \end{aligned}$$

The first term Y_i^4 , $i = m$ has N choices. Then j and n both have $N - 1$ choices. So it has $N(N - 1)^2$ terms.

The second term $Y_i^2 Y_m^2$, i has N choices. Then m has $N - 1$ choices. $j \neq i$ so j has $N - 1$ choices. $m \neq n$ so n has $N - 1$ choices. So it has $N(N - 1)^3$ terms.

The third term $Y_i^2 Y_m^2$, Since $i = m$, there are N choices. $j = n$ there are $N - 1$ choices since $j \neq i$. Therefore there are $N(N - 1)$ terms.

$$\begin{aligned} (4N^2(N-1)^2)\mathbf{E}S^4 &= 4N(N-1)^2\mathbf{E}Y_1^4 + 4N(N-1)^3\mathbf{E}Y_1^2\mathbf{E}Y_2^2 + 8N(N-1)\mathbf{E}Y_1^2\mathbf{E}Y_2^2 \\ (4N^2(N-1)^2)\mathbf{E}S^4 &= 4N(N-1)^2\theta_4 + 4N(N-1)[(N-1)^2 + 2]\theta_2^2 \\ \mathbf{E}S^4 &= \frac{1}{N}\theta_4 + \frac{N^2 - 2N + 3}{N(N-1)}\theta_2^2 \end{aligned}$$

Therefore

$$\text{Var}S^4 = \mathbf{E}S^4 - (\mathbf{E}S^2)^2 = \frac{1}{N}\theta_4 + \frac{N^2 - 2N + 3}{N(N-1)}\theta_2^2 - \theta_2^2 = \frac{1}{N} \left(\theta_4 - \frac{N-3}{N-1}\theta_2^2 \right)$$

■

(c) Let $Y_i = X_i - \theta_1$. Then $\mathbf{E}Y_i = 0$, $\mathbf{E}Y_i^j = \theta_j$

$$\begin{aligned}
\text{Cov}(\bar{X}, S^2) &= \text{Cov}(\bar{Y}, S^2) \\
&= \mathbf{E}(\bar{Y}, S^2) - \mathbf{E}\bar{Y}\mathbf{E}S^2 \\
&= \mathbf{E}(\bar{Y}, S^2) \\
&= \frac{1}{N} \frac{1}{2N(N-1)} \mathbf{E} \sum_{i \neq j, k} Y_k (Y_i - Y_j)^2 \\
&= \frac{1}{2N^2(N-1)} \mathbf{E} \sum_{i \neq j, k} Y_k Y_i^2 - 2Y_k Y_i Y_j + Y_k Y_j^2
\end{aligned}$$

Note that $Y_k Y_i Y_j$ vanishes because $i \neq j$ so the expected value of one of them will be 0. By the same argument, $Y_k Y_i^2$ will not be 0 if $k = i$. Therefore

$$\begin{aligned}
\text{Cov}(\bar{X}, S^2) &= \frac{1}{2N^2(N-1)} \mathbf{E} \sum_{i \neq j, k=i} 2Y_i^3 \\
&= \frac{1}{2N^2(N-1)} 2N(N-1) \mathbf{E}Y_i^3 \\
&= \frac{1}{N} \theta_3
\end{aligned}$$

$\text{Cov}(\bar{X}, S^2) = 0$ when $\theta_3 = 0$. ■