BIOSTAT 602 Biostatistical Inference
Homework 01

Ashton Baker

Thursday January 12, 2017

1. A coin is twice as likely to turn up tails as heads. If the coin is tossed independently, what is the probability that the third head occurs on the 5th trial?

   Solution. This implies that $P(\text{Heads}) = 1/3$ and $P(\text{Tails}) = 2/3$. The probability of 2 heads occuring in the first 4 trials is described by a binomial distribution:

$$\binom{4}{2}(1/3)^2(2/3)^2 = 8/27$$

   Then, the probability of heads occuring on the 5th trial is $1/3$, so the probability is

$$(8/27)(1/3) = 8/81$$

2. Suppose X and Y are two independent variables with unit variance. Let $Z = aX + Y$, where $a > 0$. If $\mathrm{Cor}(X, Z) = 1/3$, then obtain the value of a.

   Solution. By the definition of correlation,

$$\mathrm{Cor}(X, Z) = \frac{\mathrm{Cov}(X, Z)}{\sigma_X \sigma_Z} = 1/3$$

   We can begin by finding $\mathrm{Cov}(X, Z)$. Note that because X and Y are independent, $\mathrm{Cov}(X, Y) = 0$.

$$\begin{aligned}
\mathrm{Cov}(X, Z) &= \mathbb{E}[XZ] - \mathbb{E}[X]\,\mathbb{E}[Z] \\
&= \mathbb{E}\left[aX^2 + XY\right] - \mathbb{E}[X]\,\mathbb{E}[aX + Y] \\
&= a\mathbb{E}\left[X^2\right] + \mathbb{E}[XY] - a\mathbb{E}[X]\,\mathbb{E}[X] + \mathbb{E}[X]\,\mathbb{E}[Y] \\
&= a\mathrm{Var}(X) + \mathrm{Cov}(X, Y) \\
&= a
\end{aligned}$$

   Because X and Y have unit variance,

$$\begin{aligned}
\sigma_Z &= \sqrt{\mathrm{Var}(Z)} \\
&= \sqrt{\mathrm{Var}(aX + Y)} \\
&= \sqrt{a^2\mathrm{Var}(X) + \mathrm{Var}(Y) + 2a\mathrm{Cov}(X, Y)} \\
&= \sqrt{a^2 + 1}
\end{aligned}$$

   and $\sigma_X = 1$. So

$$\begin{aligned}
\frac{\mathrm{Cov}(X, Z)}{\sigma_X \sigma_Z} &= 1/3 \\
\frac{a}{\sqrt{a^2 + 1}} &= 1/3 \\
a &= \sqrt{a^2 + 1}/3 \\
9a^2 &= a^2 + 1 \\
a^2 &= 1/8 \\
a &= \sqrt{1/8}
\end{aligned}$$

3. Let $g(x)$, $x \geq 0$ be a valid pdf for a nonnegative random variable and define

$$f(x, y) = \frac{g(\sqrt{x^2 + y^2})}{2\pi\sqrt{x^2 + y^2}}$$

   for $x, y \in \mathbb{R}$

(a) Show that $f(x, y)$ is a valid pdf.

Solution. The function f depends on x and y only in terms of $r = \sqrt{x^2 + y^2}$. So $f(x, y) = f(r)$, and

$$\begin{aligned}
\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)\ dx\ dy &= \int_0^{\infty} 2\pi r f(r)\ dr \\
&= \int_0^{\infty} 2\pi r \frac{g(r)}{2\pi r}\ dr \\
&= \int_0^{\infty} g(r)\ dr \\
&= 1
\end{aligned}$$

So f is a valid pdf.

(b) Suppose that the pair (X, Y) has the pdf $f(x, y)$. What is $P(XY > 0)$?

Solution. The region where $XY > 0$ is the union of quadrant 1 and quadrant 3. Due to the radial symmetry of f, $P(XY > 0) = 1/2$.

4. Given independent and identically distributed random samples $X_1, X_2, \ldots, X_n$, each with finite mean μ and finite variance σ^2, define

$$\begin{aligned}
\overline{X} &= \frac{1}{n}\sum_{i=1}^{n} X_i \\
S^2 &= \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 \\
W^2 &= \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2
\end{aligned}$$

(a) Show that $S^2 \xrightarrow{P} \sigma^2$

Solution. Note that

$$\begin{aligned}
\frac{n-1}{n}S^2 &= \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}\left[(X_i - \mu) + (\mu - \overline{X})\right]^2 \\
&= \frac{1}{n}\left[\sum_{i=1}^{n}(X_i - \mu)^2 + 2(\mu - \overline{X})\sum_{i=1}^{n}(X_i - \mu) + \sum_{i=1}^{n}(\overline{X} - \mu)^2\right] \\
&= \frac{1}{n}\left[\sum_{i=1}^{n}(X_i - \mu)^2 + 2n(\mu - \overline{X})(\overline{X} - \mu) + n(\overline{X} - \mu)^2\right] \\
&= \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2 - 2(\mu - \overline{X})^2 + (\overline{X} - \mu)^2.
\end{aligned}$$

Note that $\mathbb{E}\left[(X_i - \mu)^2\right] = \sigma^2$, and $\mathbb{E}\left[\overline{X} - \mu\right] = 0$, so by the Weak Law of Large Numbers, $\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2 \xrightarrow{P} \sigma^2$ and $(X_i - \mu)^2 \xrightarrow{P} 0$. Therefore $(n-1)/nS^2 \xrightarrow{P} \sigma^2$. However, $(n-1)/n \to 1$, so $S^2 \xrightarrow{P} \sigma^2$.

(b) Derive the asymptotic distribution of $\frac{\sqrt{n}(\overline{X} - \mu)}{\sqrt{S^2}}$

Solution. Let

$$\begin{aligned}
A_n &= \frac{\overline{X}_n - \mu}{\sqrt{\sigma^2/n}} \\
B_n &= \frac{\sigma^2}{S^2}
\end{aligned}$$

By the Central Limit Theorem, $A_n \xrightarrow{d} N(0,1)$. We have proved that $S^2 \xrightarrow{d} \sigma^2$, and since B_n applies a continuous transformation to S^2, we have $B_n \xrightarrow{d} 1$. Then, by Slutsky's theorem,

$$\frac{\sqrt{n}\left(\overline{X}-\mu\right)}{\sqrt{S^2}} = A_n B_n \xrightarrow{d} N(0,1)$$

(c) Use the Delta method to derive the asymptotic distribution of $\overline{X}^2$ after you normalize it appropriately.
Solution. By the Central Limit Theorem, $\sqrt{n}(\overline{X}-\mu) \xrightarrow{d} N\left(0,\sigma^2\right)$. Then, define $g(x) = x^2$, so that $g^{(1)}(x) = 2x$ is continuous. If $\mu \neq 0$, then we can apply the Delta Method to obtain

$$\sqrt{n}\left[g(\overline{X}) - g(\mu)\right] \xrightarrow{d} N\left(0, \sigma^2\left[g^{(1)}(\mu)\right]^2\right)$$
$$\sqrt{n}\left[\overline{X}^2 - \mu^2\right] \xrightarrow{d} N\left(0, 4\sigma^2\mu^2\right)$$

5. For two sets of random varibales $\{X_i\}$, $i = 1, \ldots, n$, and $\{Y_i\}$, $j = 1, \ldots, m$, show that

$$\text{Cov}\left(\sum_{i=1}^{n} a_i X_i, \sum_{j=1}^{m} b_j Y_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{m} a_i b_j \text{Cov}\left(X_i, Y_j\right)$$

where a_i and b_j are arbitrary constants.
Solution.

$$\begin{aligned}
\text{Cov}\left(\sum_{i=1}^{n} a_i X_i, \sum_{j=1}^{m} b_j Y_j\right) &= \mathbb{E}\left[\left(\sum_{i=1}^{n} a_i X_i - \mathbb{E}\left[\sum_{i=1}^{n} a_i X_i\right]\right)\left(\sum_{j=1}^{m} b_j Y_j - \mathbb{E}\left[\sum_{j=1}^{m} b_j Y_j\right]\right)\right] \\
&= \mathbb{E}\left[\sum_{i=1}^{n} a_i \left(X_i - \mathbb{E}\left[X_i\right]\right) \sum_{j=1}^{m} b_j \left(Y_j - \mathbb{E}\left[Y_j\right]\right)\right] \\
&= \mathbb{E}\left[\sum_{i=1}^{n}\sum_{j=1}^{m} a_i b_j \left(X_i - \mathbb{E}\left[X_i\right]\right)\left(Y_j - \mathbb{E}\left[Y_j\right]\right)\right] \\
&= \sum_{i=1}^{n}\sum_{j=1}^{m} a_i b_j \mathbb{E}\left[\left(X_i - \mathbb{E}\left[X_i\right]\right)\left(Y_j - \mathbb{E}\left[Y_j\right]\right)\right] \\
&= \sum_{i=1}^{n}\sum_{j=1}^{m} a_i b_j \text{Cov}\left(X_i, Y_j\right)
\end{aligned}$$

6. Suppose $N \sim \text{Poisson}(\lambda)$. Given $N = n > 0$, $X_1, \ldots, X_N$ are iid and follow $U[0,1]$. We define $X_0 = 0$ when $N = 0$.

(a) Given $N = n$, find the probability that $X_0, X_1, \ldots, X_N$ are all less that t, where $0 < t < 1$.
Solution. Since each X_i is uniformly distributed, $P(X_i < t) = t$. If $N = n$, then because $X_1, \ldots, X_n$ are independent,

$$P(X_1, \ldots, X_n < t) = P(X_1 < t)\, P(X_2 < t) \cdots P(X_n < t) = t^n$$

(b) Find the (unconditional) probability that $X_0, X_1, \ldots, X_N$ are all less than t, where $0 < t < 1$.
Solution. In this case, we need to sum over all the possible values that N could take, weighted by the probability of observing each particular value n. So

$$\begin{aligned}
P(X_1, \ldots, X_N < t) &= \sum_{n=0}^{\infty} P(N = n)\, P(X_1, \ldots, X_n < t) \\
&= \sum_{n=0}^{\infty} \frac{\lambda^n e^{-\lambda}}{n!} t^n \\
&= e^{-\lambda(1-t)}
\end{aligned}$$

(c) Let $S_N = X_0 + X_1 + \cdots + X_N$. Compute $\mathbb{E}[S_N]$.

Solution.

$$\begin{aligned}
\mathbb{E}[S_N] &= \sum_{n=0}^{\infty} P(N = n)\, \mathbb{E}[X_1 + \cdots + X_n] \\
&= \sum_{n=0}^{\infty} \left(\frac{\lambda^n e^{-\lambda}}{n!}\right) \left(\frac{n}{2}\right) \\
&= \frac{\lambda}{2}
\end{aligned}$$