EECS 70 Discrete Mathematics and Probability Theory
Fall 2015 Jean Walrand

Discussion 14A Solution

1. (Covariance) We have a bag of 5 red and 5 blue balls. We take two balls from the bag without replacement. Let X_1 and X_2 be indicator random variables for the first and second ball being red.

What is $Cov(X_1, X_2)$?

We can use the formula $Cov(X_1, X_2) = E(X_1 X_2) - E(X_1)E(X_2)$.

$$E(X_1) = \frac{5}{10} \times 1 + \frac{5}{10} \times 0 = \frac{1}{2}$$
$$E(X_2) = \frac{5}{10} \times 1 + \frac{5}{10} \times 0 = \frac{1}{2}$$

$$E(X_1 X_2) = \frac{5}{10} \cdot \frac{4}{9} \times 1 + (1 - \frac{5}{10} \cdot \frac{4}{9}) \times 0 = \frac{2}{9}$$

Therefore,

$$E(X_1 X_2) - E(X_1)(X_2) = \frac{2}{9} - \frac{1}{2} \times \frac{1}{2} = \frac{-1}{36}$$

2. (LLSE) We have two bags of balls. The fractions of red balls and blue balls in bag A are $\frac{2}{3}$ and $\frac{1}{3}$ respectively. The fractions of red balls and blue balls in bag B are $\frac{1}{2}$ and $\frac{1}{2}$ respectively. Someone gives you one of the bags (unmarked) uniformly at random. Then we draw 6 balls from the same bag with replacement. Let X_i be the indicator random variable that ball i is red. Now, let us define $X = \sum_{1 \leq i \leq 3} X_i$ and $Y = \sum_{4 \leq i \leq 6} X_i$.

Find $LLSE(Y|X)$. [Hint: recall that $LLSE(Y|X) = E(Y) + \frac{Cov(X,Y)}{Var(X)}(X - E(X))$]

$$\begin{aligned} E(X) &= 3 \cdot E(X_1) \\ &= 3 \cdot P(X_1 = 1) \\ &= 3 \cdot (\frac{1}{2} \cdot \frac{2}{3} + \frac{1}{2} \cdot \frac{1}{2}) \\ &= \frac{7}{4} \end{aligned}$$

$$E(Y) = E(X) = \frac{7}{4}$$

$$\begin{aligned} Cov(X,Y) &= Cov(\sum_{1 \leq i \leq 3} X_i, \sum_{4 \leq j \leq 6} X_j) \\ &= 9 \cdot Cov(X_1, X_4) \\ &= 9 \cdot (E(X_1 X_4) - E(X_1) \cdot E(X_4)) \end{aligned}$$

$$
\begin{aligned}
E(X_1X_4)-E(X_1)E(X_4) &= P(X_1=1,X_4=1)-P(X_1=1)^2 \\
&= [\frac{1}{2}\cdot(\frac{2}{3})^2+\frac{1}{2}\cdot(\frac{1}{2})^2]-[\frac{1}{2}\cdot(\frac{2}{3})+\frac{1}{2}\cdot(\frac{1}{2})]^2 \\
&= \frac{1}{144}
\end{aligned}
$$

$$
\begin{aligned}
Var(X) &= Cov(\sum_{1\le i\le 3} X_i, \sum_{1\le j\le 3} X_j) \\
&= 3\cdot Var(X_1)+6\cdot Cov(X_1,X_2) \\
&= 3(E(X_1^2)-E(X_1)^2)+6\cdot\frac{1}{144} \\
&= 3(\frac{7}{12}-(\frac{7}{12})^2)+6\cdot\frac{1}{144} \\
&= \frac{111}{144}
\end{aligned}
$$

So, $LLSE(Y|X)=\frac{7}{4}+\frac{9}{111}(X-\frac{7}{4})=\frac{3}{37}X+\frac{119}{74}$

3. (Confidence interval) Let $\{X_i\}_{1\le i\le n}$ be a sequence of iid Bernoulli random variables with parameter μ. Assume we have enough samples such that $P(|\frac{1}{n}\sum_{1\le i\le n}X_i-\mu|>0.1)=0.05$.

Can you give 95% confidence interval for μ if you are given the outcomes of X_i?

$[\frac{1}{n}\sum_{1\le i\le n}X_i-0.1,\frac{1}{n}\sum_{1\le i\le n}X_i+0.1]$

4. (Chernoff's bound) Let X be a binomial random variable with parameters $(n,0.5)$.

Prove that there exists $\alpha>0$ such that $P(X>0.7n)\le e^{-\alpha n}$. [Hint : $\frac{e^{-0.7}+e^{0.3}}{2}=0.923<1$]

Let $X=\sum_{1\le i\le n}X_i$, where X_i are iid Bernoulli random variable with parameter 0.5.

Also let $t>0$.

$$
\begin{aligned}
P(X>0.7n) &= P(\sum_{1\le i\le n}X_i>0.7n) \\
&= P(e^{\sum_{1\le i\le n}tX_i}\ge e^{0.7nt}) \\
&\le \frac{E(e^{\sum_{1\le i\le n}tX_i})}{e^{0.7nt}} \\
&= \frac{\Pi_{1\le i\le n}E(e^{tX_i})}{e^{0.7nt}} \\
&= \frac{E(e^{tX_1})^n}{e^{0.7nt}} \\
&= \frac{(0.5+0.5e^t)^n}{e^{0.7nt}} \\
&= (\frac{e^{-0.7t}+e^{0.3t}}{2})^n \\
&= (0.923)^n, \text{ by taking t=1} \\
&= e^{-\ln(1/0.923)n}
\end{aligned}
$$

Since $\ln(1/0.923)>0$, we have $\alpha=\ln(1/0.923)$