Stat 134: Indicator Review

Adam Lucas

December 11th, 2019

Conceptual Review

a. What is the computational formula for Var$(X+Y)$?

   Var$(X+Y)$ = Var(X) + Var(Y) + 2Cov(X,Y)

b. How do we choose indicators?

   We look at the quantity we want. For example, in the elevator problem, the quantity we want is the expected number of floors where no one gets off, so we should choose to indicate on the floors instead of the passengers.

c. Suppose X is the sum of n identical indicators I_j's. What is Var(X)?

   $Var(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2 = \mathbb{E}(\sum_{j=1}^{n} I_j^2 + \sum_{j \neq k} I_j I_k) - [\mathbb{E}(X)]^2 = \mathbb{E}(\sum_{j=1}^{n} I_j + \sum_{j \neq k} I_j I_k) - [\mathbb{E}(X)]^2 = n\mathbb{E}(I_1) + n(n-1)\mathbb{E}(I_1 I_2) - [\mathbb{E}(X)]^2$

Problem 1

Toss a p-coin n times. Let W_r refer to the number of trials until the r_{th} head. Find $Corr(W_1, W_r)$.

Observe that $W_1 \sim Geom(p), W_r \sim NegBin(r,p)$. We can rewrite $W_r = \sum_{i=1}^{r} B_i$, where each B_i is an i.i.d. $Geom(p)$, and $B_1 = W_1$. Hence, $Cov(W_1, W_r) = Cov(B_1, \sum_{i=1}^{r} B_i) = Var(B_1) = \frac{1-p}{p^2} \implies Corr(W_1, W_r) = \frac{Cov(W_1,W_r)}{SD(W_1)SD(W_r)} = \frac{\frac{1-p}{p^2}}{\sqrt{\frac{1-p}{p^2}\frac{r(1-p)}{p^2}}} = \frac{1}{\sqrt{r}}$.

Problem 2

Suppose you order f cups of fruit tea and m cups of milk tea, along with f servings of lychee jelly and m servings of boba to add to the drinks. Ideally, you would like fruit tea with lychee jelly and milk tea with boba, but the boba shop adds one purchased topping per drink randomly. Let X be the number of ideal drinks you get in the end. Find:

a. $\mathbb{E}(X)$;

b. $\text{Var}(X)$.

a. Let I_{F_j} be the indicator that the j_{th} cup of fruit tea is ideal and I_{M_k} be the indicator that k_{th} cup of milk tea is ideal. $\implies X = \sum_{j=1}^{f} I_{F_j} + \sum_{k=1}^{m} I_{M_k}$.

$\mathbb{E}(X) = \mathbb{E}(sum_{j=1}^{f} I_{F_j} + \sum_{k=1}^{m} I_{M_k}) = f\mathbb{E}(I_{F_1}) + m\mathbb{E}(I_{M_1}) = f\frac{f}{f+m} + m\frac{m}{f+m}$.

b. Note that here the indicators are not identical, so you cannot use the derived formula from conceptual review. Instead, you will need to derive the variance formula yourself!

$$\begin{aligned}
\mathbb{E}(X^2) &= \mathbb{E}[(\sum_{j=1}^{f} I_{F_j} + \sum_{k=1}^{m} I_{M_k})^2] \\
&= \mathbb{E}[(\sum_{j=1}^{f} I_{F_j} + \sum_{k=1}^{m} I_{M_k})(\sum_{j=1}^{f} I_{F_j} + \sum_{k=1}^{m} I_{M_k})] \\
&= \mathbb{E}(\sum_{j,l\in F} I_{F_j} I_{F_l} + \sum_{k,r\in M} I_{M_k} I_{M_r} + 2\sum_{s\in F, t\in M} I_{F_s} I_{M_t}) \\
&= \mathbb{E}(\sum_{j,l\in F} I_{F_j} I_{F_l}) + \mathbb{E}(\sum_{k,r\in M} I_{M_k} I_{M_r}) + 2\mathbb{E}(\sum_{s\in F, t\in M} I_{F_s} I_{M_t}) \\
&= \mathbb{E}(\sum_{j=1}^{f} I_{F_j}^2 + \sum_{j\neq l} I_{F_j} I_{F_l}) + \mathbb{E}(\sum_{k=1}^{m} I_{M_k}^2 + \sum_{k\neq r} I_{M_k} I_{M_r}) + 2\mathbb{E}(\sum_{s\in F, t\in M} I_{F_s} I_{M_t}) \\
&= f\mathbb{E}(I_{F_1}) + f(f-1)\mathbb{E}(I_{F_1} I_{F_2}) + m\mathbb{E}(I_{M_1}) + m(m-1)\mathbb{E}(I_{M_1} I_{M_2}) + 2fm\mathbb{E}(I_{F_1} I_{M_1}) \\
&= f\frac{f}{f+m} + f(f-1)\frac{\binom{f}{2}}{\binom{f+m}{2}} + m\frac{m}{f+m} + m(m-1)\frac{\binom{m}{2}}{\binom{f+m}{2}} + 2fm\frac{f}{f+m}\frac{m}{f+m-1}
\end{aligned}$$

$$\begin{aligned}
Var(X) &= \mathbb{E}(X^2) - [\mathbb{E}(X)]^2 \\
&= f\frac{f}{f+m} + f(f-1)\frac{\binom{f}{2}}{\binom{f+m}{2}} + m\frac{m}{f+m} + m(m-1)\frac{\binom{m}{2}}{\binom{f+m}{2}} + 2fm\frac{f}{f+m}\frac{m}{f+m-1} - (f\frac{f}{f+m} + m\frac{m}{f+m})^2
\end{aligned}$$

Problem 3

A p-coin is a coin that lands heads with probability p. Flip a p-coin n times. A "run" is a maximal sequence of consecutive flips that are all the same. For example, the sequence $HTHHHTTH$ with $n = 8$ has five runs, namely H, T, HHH, TT, H. Let X denote the number of runs in these n flips. Find $\mathbb{E}(X)$.

Let I_j be the indicator that j_{th}&$(j-1)_{th}$ trials are different. The idea here is to only increment at the start of a new run.

$X = 1 + \sum_{j=2}^{n} I_j$ since the first trial is always the start of a new run.

$$
\begin{aligned}
\mathbb{E}(X) &= \mathbb{E}(1 + \sum_{j=2}^{n} I_j) \\
&= 1 + (n-1)\mathbb{E}(I_2) \\
&= 1 + (n-1)P(I_2 = 1) \\
&= 1 + (n-1)P(\text{1st trial and 2nd trial have different outcomes}) \\
&= 1 + (n-1)P(HT \text{ or } TH) \\
&= 1 + (n-1)[P(HT) + P(TH)] \\
&= 1 + (n-1)(pq + qp) \\
&= 1 + (n-1)2pq
\end{aligned}
$$