Section 13.3. Variance

Note. In this section, we define the variance of a random variable on a finite sample space and state Chebyshev's Inequality in this setting. We give an almost surely result on the value of the stability number of a random graph in $\mathcal{G}_{p,1/2}$.

Note. Recall that the variance of a random variable (finite or infinite, discrete or continuous) is $E((X-\mu)^2)$. See my online notes for Mathematical Statistics 1 on 1.9. Some Special Expectations and for Measure Theory Based Probability on 4.10. Expectation. Our definition here is the same, but since we consider finite sample spaces then expectation and things that follow from it (such as variance) are computed using finite sums.

Definition. The *variance* of random variable X on a finite sample space is

$$V(X) = E((X - E(X))^2).$$

Note. Since expectation is linear, we can also calculate the variance as

$$V(X) = E((X - E(X))^2) = E(X^2 - 2XE(X) + (E(X))^2)$$

$$= E(X^2) - 2E(X)^2 + E(X)^2 = E(X^2) - E(X)^2 = E(X^2) - E^2(X).$$

If we denote the mean of X as μ then $\mu = E(X)$ and we have $V(X) = E(X^2) - E^2(X) = E(X^2) - \mu^2$, as we see in Mathematical Statistics 1 (see Note 1.9.A of 1.9. Some Special Expectations). For X an indicator random variable (for which $X \in \{0,1\}$ and $X = X^2$) we have $E(X^2) = E(X)$ so that $V(X) = E(X) - E^2(X) \le E(X)$.

Note. In Mathematical Statistics 1, we state Chebyshev's Inequality for X a random variable (finite or infinite, discrete or continuous) where $E(X^2) < \infty$: for every $k > 0$, $P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$ (see Theorem 1.10.3 in my Mathematical Statistics 1 online notes on 1.10. Important Inequalities, and Theorem 4.10.7 in my online notes for Measure Theory Based Probability on 4.10. Expectation). Bondy and Murty describe Chebyshev's Inequality as "[it] bounds the divergence of a random variable from its mean. It plays, in some sense, a complementary role to that of Markov's Inequality." Our statement is equivalent to that given in Mathematical Statistics and the other settings, as we will show.

Theorem 13.7. Chebyshev's Inequality.
Let X be a random variable on a finite probability space and let $t > 0$. Then

$$P(|X - E(X)| \geq t) \leq \frac{V(X)}{t^2}.$$

Note. If we replace random variable X with random variable $Y = X/\sigma$ where $\sigma = \sqrt{V(X)}$ (i.e., σ is the standard deviation of X) then by the linearity of expectation we have $E(Y) = E(X/\sigma) = E(X)/\sigma$ and

$$V(Y) = V(X/\sigma) = E((X/\sigma - E(X/\sigma))^2) = E((X - E(X))^2)/\sigma^2$$

$$= V(X)/\sigma^2 = \sigma^2/\sigma^2 = 1$$

then Theorem 13.7 gives $P(|Y - E(Y)| \geq t) \leq \frac{V(Y)}{t^2}$ or $P(|X/\sigma - E(X)/\sigma| \geq t) \leq \frac{1}{t^2}$ or $P(|X - E(x)| \geq t\sigma) \leq \frac{1}{t^2}$, which is the same version of Chebyshev's Inequality given in the other sources (replacing $t > 0$ here with $k > 0$).

Note. We will use the following form of Chebyshev's Inequality in establishing "almost surely" results.

Corollary 13.8. Let X_n be a random variable in a finite probability space (Ω_n, P_n) where $n \geq 1$. If $E(X_n) \neq 0$ and $V(X_n) \ll E^2(X_n)$, then $P(X_n = 0) \to 0$ as $n \to \infty$.

Definition. The *covariance* $C(X,Y)$ of two random variables X and Y on a finite probability space is $C(X,Y) = E(XY) - E(X)E(Y)$.

Note 13.3.A. Let $G \in \mathcal{G}_{n,p}$ be a random graph and let X be the random variable representing the number of triangles in G. In Note 13.2.B we saw that if $pn \to 0$ then G is almost surely triangle-free. We now show that if $pn \to \infty$ then G almost surely has at least one triangle. For $S \subseteq V$ with $|S| = 3$, let A_S be the event that the induced subgraph, $G[S]$, is a triangle and let X_S be the indicator random variable for A_S. Then $X = \sum_{S \subseteq V, |S|=3} X_S$. As shown in Note 13.2.B, $E(X) = \binom{n}{3} p^3$. It is to be shown in Exercise 13.3.1 that

$$V(X) \leq E(X) + \sum_{S \subseteq V, |S|=3, S \neq T} C(X_S, X_T). \qquad (*)$$

If $|S \cap T| \in \{0,1\}$ then $G[S]$ and $G[T]$ can have no common edges, so $E(X_S X_T) = p^6 = E(X_S)E(X_T)$ and hence the covariance is $C(X_S, X_T) = 0$. If $|S \cap T| = 2$ then $G[S]$ and $G[T]$ have one potential edge in common, so $C(X_S, X_T) = E(X_S X_T) - E(X_S)E(X_Y) = p^5 - p^6$ (since X_S and X_T are indicator random variables then $X_S X_T = 1$ only when $G[S]$ includes each of the three relevant edges <u>and</u> $G[T]$ contains the other two relevant edges [i.e., those two edges not in $G[S]$], and otherwise $X_S X_T = 0$ so that $E(X_S X_T) = p^5$). Now we can choose the elements of S and T where $|S \cap T| = 2$ by first choosing the two distinct elements of $S \cap T$ in any way (which can be done in $\binom{n}{2}$ ways), then choosing the third element of S (which is distinct from the two elements of $S \cap T$ and so can be done in $n-2$ ways), and finally choosing the third element of T (which is distinct from the three elements already chosen, which can be done in $n-3$ ways). So there are $\displaystyle\binom{n}{2}(n-2)(n-3)$ ways to choose a pair of sets $S, T \subseteq V$ with $|S| = |T| = 3$, and $|S \cap T| = 2$. So by $(*)$

$$\begin{aligned}
V(X) &\le E(X) + \sum_{S\subseteq V, |S|=3, S\neq T} C(X_S, X_T) \\
&= E(X) + \sum_{S\subseteq V, |S|=3, |S\cap T|=0} C(X_S, X_T) + \sum_{S\subseteq V, |S|=3, |S\cap T|=1} C(X_S, X_T) \\
&\quad + \sum_{S\subseteq V, |S|=3, |S\cap T|=2} C(X_S, X_T) \\
&= \binom{n}{3}p^3 + 0 + 0 + \binom{n}{2}(n-2)(n-3)(p^5 - p^6) \\
&\le \binom{n}{3}p^3 + \binom{n}{2}(n-2)(n-3)p^5.
\end{aligned}$$

Since $E(X) = \binom{n}{3}p^3$ then

$$\begin{aligned}
\frac{V(X)}{E^2(X)} &\le \frac{1}{\binom{n}{3}p^3} + \frac{\binom{n}{2}(n-2)(n-3)p^5}{\left(\binom{n}{3}p^3\right)^2} \\
&= \frac{6}{n(n-1)(n-2)p^3} + \frac{36n(n-1)(n-2)(n-3)p^5}{2n^2(n-1)^2(n-2)^2p^6} \\
&= \frac{6}{n(n-1)(n-2)p^3} + \frac{18(n-3)}{n(n-1)(n-2)p} \\
&< \frac{6}{n(n/2)(n/2)p^3} + \frac{18}{n(n/2)p} \text{ since } n-1 > n/2,\ n-2 > n/2,
\end{aligned}$$

$$\text{and } \frac{n-3}{n-2} < 1 \text{ for } n \geq 3$$
$$= \frac{24}{(np)^3} + \frac{36}{n(np)}.$$

Now if $pn \to \infty$ then $n \to \infty$ (and $p \neq 0$) since $p \leq 1$. So if $pn \to \infty$ then $V(X)/E^2(X) \to 0$ as $n \to \infty$; that is, $V(X) \ll E^2(X)$. So (replacing X here with X_n) by Corollary 13.8, $P(X = 0) \to 0$ as $n \to \infty$. That is, if $pn \to \infty$ as $n \to \infty$ the G almost surely has at least one triangle.

Note. In Theorem 13.6, we proved that for G a random graph in $\mathcal{G}_{n,p}$, the stability number $\alpha(G)$ is almost surely at most $\lceil 2p^{-1} \log n \rceil$. If $p = 1/2$, it is to be shown in Exercise 13.2.11 that this bound can be refined to $\lceil 2 \log_2 n \rceil$. Considering $p = 1/2$ again, we use Corollary 13.8 to give a sharper "almost surely" bound on $\alpha(G)$. The result is originally due to Bela Bollobás and Paul Erdös (published in *Mathematical Proceedings of the Cambridge Philosophical Society*) and, independently, D. Matula in 1976.

Theorem 13.9. Let $G \in \mathcal{G}_{n,1/2}$. For $0 \leq k \leq n$, set $f(k) = \binom{n}{k} 2^{-\binom{k}{2}}$ and let k^* be the least value of k for which $f(k)$ is less than one. Then almost surely the stability number of G, $\alpha(G)$, takes one of the three values $k^* - 2$, $k^* - 1$, or k^*.

Note. Theorem 13.9 can be refined under hypotheses concerning the rate of growth of $f(k^*)$ as follows; the proof is to be given in Exercise 13.3.2.

Corollary 13.10. Let $G \in \mathcal{G}_{n,1/2}$, and let f and k^* be as defined in Theorem 13.9. Then either:

1. $f(k^*) \ll 1$, in which case almost surely $\alpha(G)$ is equal to either $k^* - 2$ or $k^* - 1$, or
2. $f(k^* - 1) \gg 1$, in which case almost surely $\alpha(G)$ is equal to either $k^* - 1$ or k^*.

Note. Of course, it follows from Corollary 13.10 that if both $f(k^*) \ll 1$ and $f(k^* - 1) \gg 1$ then almost surely $\alpha(G) = k^* - 1$.

Revised: 3/24/2021