Chris Piech
CS109

Handout #27
April 29th, 2016

# Variance from Events

## Expectation

As we left off, we had shown that $E[\sum_i X_i] = \sum_i E[X_i]$. We are going to use that to show how to calculate variances.

### Expectation of Binomial

First let's start with some practice with the sum of expectations of indicator variables. Let $Y \sim Bin(n, p)$, in other words if $Y$ is a Binomial random variable. We can express $Y$ as the sum of $n$ Bernoulli random indicator variables $X_i \sim Ber(p)$. Since $X_i$ is a Bernoulli, $E[X_i] = p$

$$Y = X_1 + X_2 + \cdots + X_n = \sum_{i=1}^{n} X_i$$

Let's formally calculate the expectation of $Y$:

$$\begin{aligned}
E[Y] &= E[\sum_{i}^{n} X_i] \\
&= \sum_{i}^{n} E[X_i] \\
&= E[X_0] + E[X_1] + \ldots E[X_n] \\
&= np
\end{aligned}$$

### Expectation of Negative Binomial

Recall that a Negative Binomial is a random variable that semantically represents the number of trials until $r$ successes. Let $Y \sim NegBin(r, p)$.

Let $X_i$ = # trials to get success after ($i$–1)st success. We can then think of each $X_i$ as a Geometric RV: $X_i \sim Geo(p)$. Thus, $E[X_i] = \frac{1}{p}$. We can express $Y$ as:

$$Y = X_1 + X_2 + \cdots + X_r = \sum_{i=1}^{r} X_i$$

Let's formally calculate the expectation of $Y$:

$$\begin{aligned}
E[Y] &= E[\sum_{i=1}^{r} X_i] \\
&= \sum_{i=1}^{r} E[X_i] \\
&= E[X_1] + E[X_2] + \ldots E[X_r] \\
&= \frac{r}{p}
\end{aligned}$$

## Variance

Recall that $I_i$ is an indicator variable for event $A_i$ when:

$$I_i = \begin{cases} 1 & \text{if } A_i \text{ occurs} \\ 0 & \text{otherwise } n \equiv 1 \end{cases}$$

Let $X$ be the # of events that occur: $X = \sum_{i=1}^{n} I_i$

$$E[X] = E[\sum_{i=1}^{n} I_i] = \sum_{i=1}^{n} E[X_i] = \sum_{i=1}^{n} P(A_i)$$

Now consider a pair of events $A_iA_j$ occurring. $I_iI_j = 1$ if both events $A_i$ and $A_j$ occur, 0 otherwise. The number of pairs of events that occur is $\binom{X}{2} = \sum_{i<j} I_iI_j$. We can write expressions for the expectation of both sides of this equation:

$$E\left[\binom{X}{2}\right] = E\left[\frac{X(X-1)}{2}\right] = \frac{1}{2}(E[X^2] - E[X])$$

$$E\left[\sum_{i<j} I_iI_j\right] = \sum_{i<j} E[I_iI_j] = \sum_{i<j} P(A_iA_j)$$

By transitivity of equality it must hold that:

$$\frac{1}{2}(E[X^2] - E[X]) = \sum_{i<j} P(A_iA_j)$$

$$E[X^2] - E[X] = 2\sum_{i<j} P(A_iA_j) \qquad E[X^2] = 2\sum_{i<j} P(A_iA_j) + E[X]$$

This gives us a way to calculate the second moment for sets of events! Now we can plug it into the definition of variance:

$$\begin{aligned} Var(X) &= E[X^2] - (E[X])^2 \\ &= 2\sum_{i<j} P(A_iA_j) + E[X] - (E[X])^2 \end{aligned}$$

### Variance of Binomial Proof

A Binomial is the sum of $n$ Bernoulli random variables: $X_i \sim Ber(p)$. Since each Bernoulli variable is an indicator we know that $E[X_i] = p$ and we can use our new way of calculating variance from events to find the variance of our Binomial.

First we must calculate $\sum_{i<j} P(A_iA_j)$

$$\begin{aligned} \sum_{i<j} P(A_iA_j) &= \sum_{i<j} p^2 && \text{since the events are independent in a Binomial} \\ &= \binom{n}{2} p^2 && \text{since the inner term of the sum is independent of i and j} \quad = \frac{n(n-1)}{2}p^2 \end{aligned}$$

Which we can plug into our new expression for variance.

$$\begin{aligned} \mathrm{Var}(X) &= 2\frac{n(n-1)}{2}p^2 + np - (np)^2 \\ &= n^2p^2 - np^2 + np - n^2p^2 \\ &= np - np^2 \\ &= np(1-p) \end{aligned}$$