Variance

Variance

Definition 4.6.1 (Variance and standard deviation). The **variance** of an r.v. X is Var$(X) = E(X - EX)^2$
Square root of the variance is called the **standard deviation** (SD):
SD$(X) = \sqrt{\text{Var}(X)}$

Recall that by $E(X - EX)^2$ we mean expectation of r.v. $(X - EX)^2$, ***not*** $(E(X - EX))^2$ (which is 0 by linearity).

Variance measures the “spread” – hence squaring the deviation from the mean, $X - EX$, (so negative deviations don’t cancel positive ones).
Why not take $E(|X - EX|)$? This measure of “spread” also exists (called MAD) but it’s not so nice, since $|x|$ is not diff. at $x = 0$.

Variance

Theorem 4.6.2 For any r.v. X, $\text{Var}(X) = E(X^2) - (EX)^2$

Proof: Let $\mu = EX$. Expanding $(X - \mu)^2$ and using linearity,

$$E(X-\mu)^2 = E(X^2 - 2\mu X + \mu^2) = E(X^2) - 2\mu EX + \mu^2 = E(X^2) - \mu^2$$

Variance has the following properties:

1) $\text{Var}(X + c) = \text{Var}(X)$ for any constant c

2) $\text{Var}(cX) = c^2\,\text{Var}(X)$ for any constant c (Var is ***not*** linear!)

3) If X and Y – independent, then $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$

In general, there's an inequality (≥), in the extreme case $Y = X$:

$\text{Var}(X + Y) = \text{Var}(2X) = 4\text{Var}(X) > 2\text{Var}(X) = \text{Var}(X) + \text{Var}(Y)$

4) $\text{Var}(X) \geq 0$, with $=$ iff $P(X = a) = 1$ for some constant a

Variance

Example 4.6.4 (Geometric and Negative Binomial). $X \sim$ Geom(p) – we know $E(X) = q/p$. By LOTUS,

$$E(X^2) = \sum_{k=0}^{\infty} k^2 P(X = k) = \sum_{k=0}^{\infty} k^2 pq^k = \sum_{k=1}^{\infty} k^2 pq^k = ?$$

Let's use the trick with differentiating $\sum_{k=0}^{\infty} q^k = \frac{1}{1-q}$ w.r.t. q:

$\sum kq^{k-1} = \frac{1}{(1-q)^2}$ – now $\cdot\, q$, so $\sum kq^k = \frac{q}{(1-q)^2}$, diff. again:

$\sum k^2 q^{k-1} = \frac{1+q}{(1-q)^3}$, so $E(X^2) = pq\frac{1+q}{(1-q)^3} = \frac{q(1+q)}{p^2}$

Finally, Var$(X) = E(X^2) - (EX)^2 = q(1+q)/p^2 - (q/p)^2 = q/p^2$

For $Y \sim$ NBin(r, p), $Y =$ (Σ of r i.i.d. Geom(p)), so Var$(Y) = rq/p^2$

Variance

Example 4.6.5 (Binomial variance). $X \sim \text{Bin}(n, p)$. With $X = I_1 + I_2 + \ldots + I_n$, $I_k =$ (indicator of k-th trial being success).

Each $\text{Var}(I_k) = E(I_k^2) - (EI_k)^2 = p - p^2 = p(1-p)$

Since all I_k are ***independent***,

$\text{Var}(X) = \text{Var}(I_1) + \ldots + \text{Var}(I_n) = np(1-p)$

Alternatively, one can find $E(X^2)$ by first finding $E($ # of ***pairs*** of successful trials): with I_k for each, $E\binom{X}{2} = \binom{n}{2}p^2$

So $n(n-1)p^2 = E(X(X-1)) = E(X^2) - E(X) = E(X^2) - np$

and $\text{Var}(X) = E(X^2) - (EX)^2 = \ldots = np(1-p)$