Variance and Standard Deviation

The expected value tells us something about the distribution but not how spread out it is!

$(10, 10, 10, 10, 10, 10, 10)$
$(0, 0, 0, 70, 0, 0, 0)$ } same mean, different variance

The variance of a random variable X, VX

$$VX = E((X - EX)^2)$$

The square eliminates the effect of sign on the differences

The standard deviation σ, $\sigma = \sqrt{VX}$

Since, $VX = E((X - EX)^2) = E(X^2 - 2X(EX) + (EX)^2)$

distribute outer E → $= E(X^2) - 2E(X)E(X) + E(X) \cdot E(X)$

eliminate on constants

$$\boxed{VX = E(X^2) - E(X)^2}$$

If X & Y are independent,

$V(X+Y) = E((X+Y)^2) - E(X+Y)^2 = E(X^2 + Y^2 + 2XY) - (EX + EY)^2$

$=$ after simplification $= VX + VY$

An Application: Strategy + the Designated Hitter

In building a baseball team, you seek batters who are good at hitting baseballs and pitchers who are good at stopping them. Pitchers are usually not good at hitting.

The National League makes the pitcher bat, while the American League lets you use a "Designated Hitter" to bat for them.

Many people don't like the Designated Hitter, claiming it takes away "strategy" by requiring the manager to make fewer pinch hitting decisions.

A "pinch hitter" is a batter who replaces someone for the rest of the game, and is usually just brought in to hit.

The DH rule certainly cuts down on the use of pinch hitters. But does it cut down on strategy?

If we take the mean and standard deviation of pinch hitter use for each team in the league, we can determine how the use of pinch hitters varies from team to team.

Source: The Bill James Historical Baseball Abstract

STANDARD DEVIATIONS OF PINCH HITTERS USED 1968–86

	American League		National League	
Year	Average	Standard Deviation	Average	Standard Deviation
1968	204	36.51	176	19.99
1969	205	36.92	190	26.42
1970	212	40.53	201	39.87
1971	190	37.09	188	24.10
1972	182	24.26	173	35.56
1973	(data not available)		202	30.87
1974	105	✓26.42	220	25.46
1975	99	✓23.95	219	23.66
1976	112	✓22.10	213	21.80
1977	116	✓31.12	228	26.69
1978	112	✓39.00	201	29.56
1979	116	✓51.97	205	23.77
1980	134	✓41.24	217	40.73
1981	86	✓24.39	153	22.67
1982	124	✓49.30	222	31.00
1983	126	✓40.14	223	29.06
1984	148	41.40	✓272	41.74
1985	141	✓42.61	274	27.70
1986	121	✓44.97	287	37.72

In all but one season, the standard deviation was higher in the American League, thus showing the managers in the AL display a wider variety of strategies on pinch hitter use than the NL.

The Chebyshev Inequality

Although just knowing the mean didn't tell us much about the distribution, the mean + variance together do!

$$VX = \sum_{w \in \Omega} (X(w) - EX)^2 P_r(w)$$

$$\geq \sum_{\substack{w \in \Omega \\ (X(w)-EX)^2 \geq \alpha}} (X(w) - EX)^2 P_r(w)$$

sum over a subset of the elements

$$\geq \sum_{\substack{w \in \Omega \\ (X(w)-EX)^2 \geq \alpha}} \alpha P_r(w) = \alpha \cdot P_r((X-EX)^2 \geq \alpha)$$

substitute smaller quantity

Thus: $P_r((X-EX)^2 \geq \alpha) \leq VX/\alpha$, for all $\alpha > 0$

Substituting $\alpha = c^2 VX$, we get

$$P_r((X-EX)^2 \geq c^2 VX) \leq 1/c^2$$

and taking the square root inside we get

$$\boxed{P_r(|X - EX| \geq c\sigma) \leq 1/c^2}$$

* Meaning X lies more than c standard deviations away from the mean with probability at most $1/c^2$.

Mean and Variance

The Expected Value of a Random Variable is

$$EX = \sum_{x \in X(\Omega)} x \cdot P_r(X=x) = \sum_{\omega \in \Omega} X(\omega)\, P_r(\omega)$$

The variance of a random variable is

$$VX = E\left((X-EX)^2\right)$$
$$= E(X^2) - (EX)^2$$

Observe that these formulae are computations on a complete probability space.

Determining Mean + Variance from Empirical Data

Suppose we obtain independent empirical observations $X_1, X_2 \cdots X_N$ for some phenomenon, what can we guess about the underlying distribution?

Estimated Mean $\hat{E}_X = \dfrac{X_1 + X_2 \cdots X_N}{N}$

Estimated Variance $\hat{V}_X = \dfrac{X_1^2 + X_2^2 + \cdots X_N^2}{N-1} - \dfrac{(X_1 + X_2 + \cdots X_N)^2}{N(N-1)}$

But why do we use $N-1$? If we had the actual distribution and each outcome was equally likely, we would just divide by N. But we don't and it isn't!

The advantage of this formulation is:

$$E(\hat{V}_X) = VX$$

Note X_k is the random variable giving the value of the k^{th} observation. These observations are assumed independent.

And we can also make an estimate of the variance, using the formula

$$\hat{V}X = \frac{X_1^2 + X_2^2 + \cdots + X_n^2}{n-1} - \frac{(X_1 + X_2 + \cdots + X_n)^2}{n(n-1)}. \tag{8.20}$$

The $(n-1)$'s in this formula look like typographic errors; it seems they should be n's, as in (8.19), because the true variance VX is defined by expected values in (8.15). Yet we get a better estimate with $n-1$ instead of n here, because definition (8.20) implies that

$$E(\hat{V}X) = VX. \tag{8.21}$$

Here's why:

$$\begin{aligned}
E(\hat{V}X) &= \frac{1}{n-1} E\Bigl(\sum_{k=1}^{n} X_k^2 - \frac{1}{n}\sum_{j=1}^{n}\sum_{k=1}^{n} X_j X_k\Bigr) && \Big\}\ \text{factor out } N-1 \\
&= \frac{1}{n-1}\Bigl(\underbrace{\sum_{k=1}^{n} E(X_k^2)}_{\text{independently drawn from dist}} - \frac{1}{n}\sum_{j=1}^{n}\sum_{k=1}^{n} E(X_j X_k)\Bigr) && \Big\}\ \text{distribute the expected value} \\
&= \frac{1}{n-1}\Bigl(\sum_{k=1}^{n} E(X^2) - \frac{1}{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\bigl(\underbrace{E(X)^2[j \neq k]}_{\text{independent}} + \underbrace{E(X^2)[j = k]}_{\text{not independent!}}\bigr)\Bigr) \\
&= \frac{1}{n-1}\Bigl(nE(X^2) - \frac{1}{n}\bigl(nE(X^2) + n(n-1)E(X)^2\bigr)\Bigr) \\
&= E(X^2) - E(X)^2 = VX.
\end{aligned}$$

And thats why the $N-1$ is there!

Note: $E(x)^2 \neq E(x^2)$

$X = \begin{array}{l} P_r(1) = 1/4 \\ P_r(2) = 1/2 \\ P_r(3) = 1/4 \end{array}$

$E(x) = \frac{1}{4} + \frac{2}{2} + \frac{3}{4} = 2$

$E(x^2) = \frac{1}{4} + \frac{4}{2} + \frac{9}{4} = \frac{18}{4}$

$E(x)^2 = \cancel{2}\ 4$

$E(x)^2 \neq E(x^2)$

"Linda is 30, a single woman with an advanced degree. While at college, she was active in protesting against nucler power"

Rank the following statements about Linda in order of decreasing likelyhood.

---

1. Linda has blond hair
2. Linda has dark hair
3. Linda is a feminist
4. Linda works as a bank teller
5. Linda works in management
6. Linda is a bank teller + a feminist
7. Linda is a blond + a bank teller
8. Linda is a blond + works in mangemt.

Derangements

Suppose a secretary randomly stuffs N letters into N addressed envelopes. What is the probability none of them end up where they are supposed to?

A derangement is a permutation σ such that $\sigma_i \neq i$ for any element - in other words no element is "where it should be".

The number of derangements is given by

$$D_N = (N-1)\, D_{N-1} + (N-1)\, D_{N-2},$$

since for any derangement, when we swap the position containing N with the last position, the first $N-1$ element either form a derangement or form a derangement + a fixed point. In each case, there are $N-1$ ways to swap.

$$\sigma_1 \; \sigma_2 \; \sigma_3 \; N \; . \; \sigma_{N-1} \; \sigma_N$$

In fact, $D_N / N! \to 1/e$, so the probability is essentially independent of N!

Fixed Points

A fixed point in a permutation π is an element $\pi_i = i$. Thus a derangement has no fixed point. What is the probability a random permutation of N elements has exactly k fixed points?

Once K fixed points have been chosen, the rest of the permutation is a derangement of size $(N-k)$, so

$h(N,k)$ = # of permutations of size N with k fixed points

$$= \binom{N}{k} D(N-k)$$

Summing over all values of k gives an alternate recurrence for $D(N)$

$$N! = \sum_k h(N,k) = \sum_k \binom{N}{k} D(N-k)$$

variable substitution and $\binom{N}{k} = \binom{N}{N-k}$

$$\boxed{N! = \sum_k \binom{N}{k} D(k)}$$

But what is the mean/variance of the number of fixed points in a random permutation? F_N is a random variable.

$$F_N(\pi) = F_{N,1}(\pi) + F_{N,2}(\pi) + \cdots \quad F_{N,N}(\pi)$$

$$F_{N,k}(\pi) = (\text{position } k \text{ of } \pi \text{ is a fixed point})$$

π is a permutation of length N.

Clearly, $F_{N,i} = F_{N,j}$, so $E(F_N) = N\,E(F_{N,1})$

Since $(N-1)!$ permutations have 1 in the first place,

$$E(F_N) = \frac{N \cdot (N-1)!}{N!} = 1$$

The expected number of fixed points per permutation is 1.

Since $V(F_N) = E(F_N^2) - (E(F_N))^2$
$= E(F_N^2) - 1$,

we need to know expected value of the squares of the fixed points.

$$E(F_N^2) = E\left(\left(\sum_{k=1}^{N} F_{N,k}\right)^2\right)$$

we cannot just move the Expected Value inside the summation because $F_{N,i}$ is not independent of $F_{N,j}$

$$= E\left(\sum_{j=1}^{N}\sum_{k=1}^{N} F_{Nj}F_{Nk}\right)$$ N^2 terms in the product

$$= \sum_{j=1}^{N}\sum_{k=1}^{N} E(F_{Nj}F_{Nk})$$ $E(X+Y) = E(X) + E(Y)$ for all random variables

$$= \sum_{1\le k\le N} E(F_{N,k}^2) + 2\sum_{1\le j<k\le N} E(F_{Nj}F_{Nk})$$

$=0$ or 1 so $F_{Nk}^2 = F_{Nk}$

$E(F_{Nj}F_{Nk}) = \sum 1 \cdot P_r(i \,\&\, j \text{ are fixed point in } \pi)$

$= 2\binom{N}{2}\left(\frac{(N-2)!}{N!}\right)$

$$= N \cdot \frac{1}{N} + 2\binom{N}{2}\frac{1}{N(N-1)}$$

$$= 1 + \frac{2N(N-1)}{2(N-1)N} = 2$$

So $V(F_N) = 2 - 1 = 1$ and the number of fixed points is 1 ± 1.

Probability Generating Functions

We have seen how generating functions make working with sequences of numbers easier. An integer valued random variable can be thought to define a sequence of probabilities....

$$G_X(z) = \sum_{k \geq 0} P_r(X=k) z^k$$

As with our other applications of generating functions, this is an alternate notation containing all information about random variable X.

Since a random variable is a function over all elementary events in a probability space...

$$G_X(z) = \sum_{\omega \in \Omega} P_r(\omega) z^{X(\omega)}$$

$$= E(z^X).$$

Since the coefficients of $G_X(z)$ are non-negative and sum to 1, $G_X(1) = 1$

Means and Variances of PGFs!

We can use Probability generating functions to compute expected values + variances in interesting ways!

Since

$$EX = \sum_{k \geq 0} k \cdot P_r(X=k)$$

and

$$G_X(z) = \sum_{k \geq 0} P_r(X=k) z^k,$$

taking the derivative of $G_X(z)$

$$G'_X(z) = \sum_{k \geq 0} k \; P_r(X=k) z^{k-1}$$

and thus: $EX = G'_X(1)$ for any random variable. This is an example of using a GF as a formula.

Since the variance, $VX = E(X^2) - (EX)^2$, to use generating functions to compute it we need $E(X^2)$.

$$E(X^2) = \sum_{k \geq 0} k^2 \cdot P_r(X=k)$$

Since,

$$G_X'(z) = \sum_{k \geq 0} k \cdot P_r(X=k)\, z^{k-1}$$

$$G_X''(z) = \sum_{k \geq 0} k(k-1)\, P_r(X=k)\, z^{k-2}$$

Therefore $E(X^2) = G_X''(1) + G_X'(1)$

and $\boxed{VX = G_X''(1) + G_X'(1) - \left(G_X'(1)\right)^2}$

Certain probability density functions have a nice form - for example, the uniform distribution of order N:

$\{0, 1, 2 \ldots, N-1\}$, $P_r(x=k) = 1/N$

$$U_N(z) = \frac{1}{N} + \frac{z}{N} + \frac{z^2}{N} \cdots \frac{z^{N-1}}{N} = \frac{1}{N}\left(\frac{1-z^N}{1-z}\right), \quad N \geq 1$$

Fixed Points via Probabilistic Gen. Functions

We seek the Expected value + variance of the random variable - How many fixed points in a permutation of size N?

Earlier, we showed $F_{N,k} = \binom{N}{k} D(N-k)$, so the p.g.f

$$\sum_k (F_N = k) z^k = \sum_k \binom{N}{k} \frac{D(N-k)}{N!} z^k$$

So

$$F_N(z) = \sum_k \frac{D(N-k)}{k!(N-k)!} z^k$$

Since

$$E(F_N) = F_N'(1) \quad \text{and} \quad VX(F_N) = F''(1) + F'(1) - F'(1)^2$$

we seek derivatives.

$$F_N'(z) = \sum_k \frac{D(N-k)}{(N-k)!} \frac{z^{k-1}}{(k-1)!} = \sum_k \frac{D(N-k-1)}{(N-k-1)!} \frac{z^k}{k!}$$

$$= F_{N-1}(z), \text{ which is a p.g.f.}$$

Thus $F_N(1) = F_{N-1}(1) = F_N''(1) = 1$

and $E(F_N) = 1 \qquad VX(F_N) = 1 + 1 - 1 = 1$

We have seen that multiplying generating functions convolve the terms. What does this mean for probability generating functions?

$$P_r(X+Y=n) = \sum_k P_r(X=k \text{ and } Y=n-k)$$

when $X+Y$ are independent random variables

$$P_r(X+Y=n) = \sum P_r(X=k)\, P_r(Y=n-k)$$

Thus multiplying the probability generating functions of two independent random variables gives the pgf of the sum of the random variables.

Example: Dice

$$G(z) = \frac{z+z^2+z^3+z^4+z^5+z^6}{6}$$ number of spots on fair die

$$EX_G = G'(1) = \frac{1+2+3+4+5+6}{6} = \frac{7}{2}$$

The number of spots on two indedent dice $= G^2(z)$

$$G^2(z) = \frac{z^2+2z^3+3z^4+4z^5+5z^6+6z^7+5z^8+4z^9+3z^{10}+2z^{11}+z^{12}}{36}$$

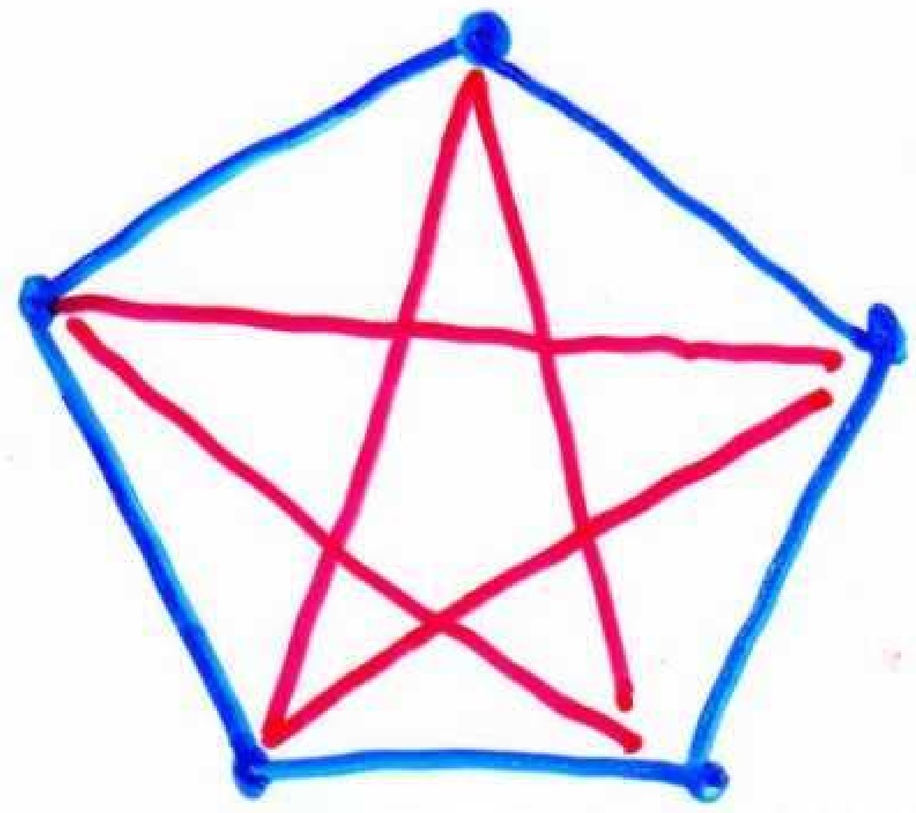
$$EX_{G^2} = \frac{2+6+12+20+30+42+40+36+30+22+12}{36}$$

The Probabilistic Method in Combinatorics

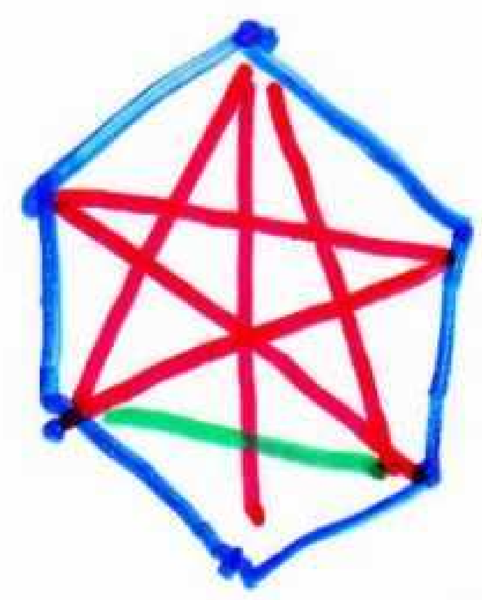
In certain problems, it can be very hard to specify an actual solution, but not as hard to prove that a (or many) solutions exist. The probabilistic method uses probability arguments to show that certain objects definitely exist!

Ramsey Numbers

Suppose we define a graph on N people, with an edge between two people who know each other. For $N \leq 5$, there are not necessarily three people, who either all know each other or are complete strangers...



However, for $N \geq 6$, no matter how you color a graph red and blue, you get at least one of them.



The Ramsey number $R(k,k)$ is defined as the smallest N such that any red-blue coloring of K_N contains a mono-chromatic K_k.

$R(3,3) = 6$, $R(4,4) = 17$, $42 \leq R(5,5) \leq 55$

Finding exact values for Ramsey numbers is notoriously difficult (Graham's licence plate reads RAMSEY !) but we can use a simple argument to get a decent lower bound.

Suppose I take K_N and flip a coin to decide whether each edge is red or blue. The probability that a given k-subset is monochromatic is $\dfrac{2}{2^{\binom{k}{2}}}$

All told, there are $\binom{N}{k}$ k-subsets of vertices. To find the probability that none of them is monochromatic is difficult, however since

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$$
$$\leq P(A) + P(B)$$

The probability of a monochromatic ~~triangle~~ k-subset is at most $\binom{N}{k} \dfrac{2}{2^{\binom{k}{2}}}$.

therefore, the probability that there does not exist a monochromatic K_k is at least

$$1 - \binom{N}{k} \frac{1}{2^{\binom{k}{2}-1}}$$

whenever $\binom{N}{k} \frac{1}{2^{\binom{k}{2}-1}} < 1$, the probability is non-zero, so there must exist a coloring without such a clique and $\boxed{R(k,k) \geq N}$!

Since $\binom{N}{k} \approx N^k$ and $2^{1-\binom{k}{2}} \approx 2^{-k^2/2}$

and $N^k / 2^{k^2/2} \sim \underline{1} \rightarrow N \approx \left(2^{k^2/2}\right)^{1/k}$

$$\boxed{R(k,k) \gtrsim 2^{k/2}}$$

Tighter bounds follow from analysis using Stirling's formula