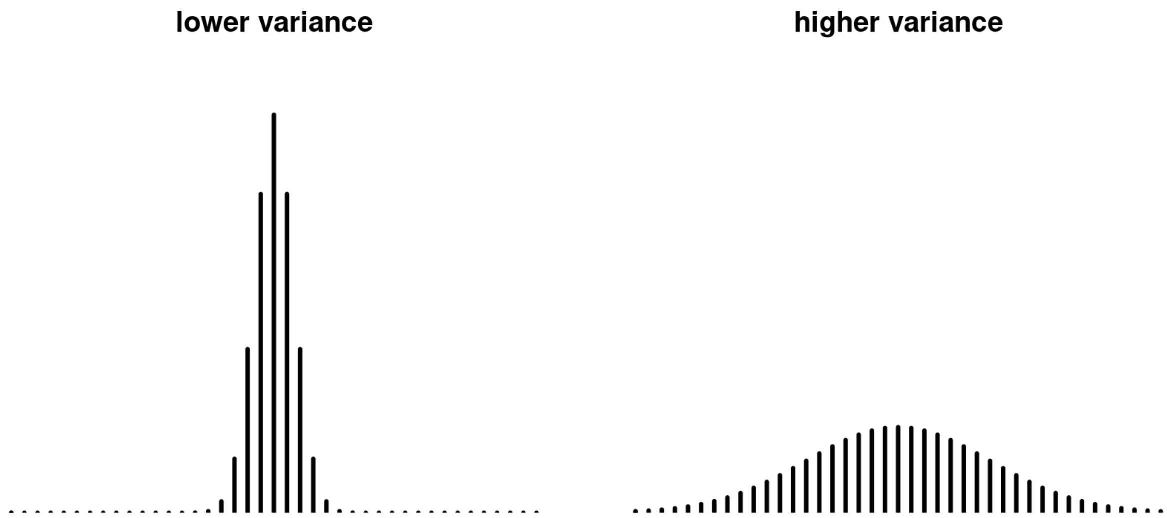

Figure 3.9

3.3.3 Variance

The expected value $E(X)$ of a random variable is a convenient single number summary that describes the "average" or "typical" behavior of a random variable. It captures the location or central tendency of the distribution. Here is another summary that describes the spread of the distribution:

Definition 3.11. Let X be *any* random variable with finite mean. Then the **variance** of X is

$$\text{var}(X) = E\left[(X - E(X))^2\right]. \tag{3.10}$$

The **standard deviation** of X is the square root of the variance: $\text{sd}(X) = \sqrt{\text{var}(X)}$.

Recalling LOTUS, we see that the variance is simply the expected value of a particular transformation of X: $g(X) = (X - E(X))^2$. This transformation g is the squared distance between X and its mean, and so the variance answers the question "how far away is X from its mean, on average?" If the answer is "not that far," then the distribution of X is not that spread out. If the answer is "pretty far," then the distribution is more spread out. Figure 3.9 displays a cartoon of this. So the variance is a single number ranging from 0 to ∞ that summarizes how variable or surprising X is. If X has low variance near 0, you don't expect to be surprised by its realizations. It's basically giving you results close to the mean. If X has high variance, you expect to regularly be surprised by what it delivers. $E(X)$ may be the typical value, but values quite far away from $E(X)$ remain fair game.

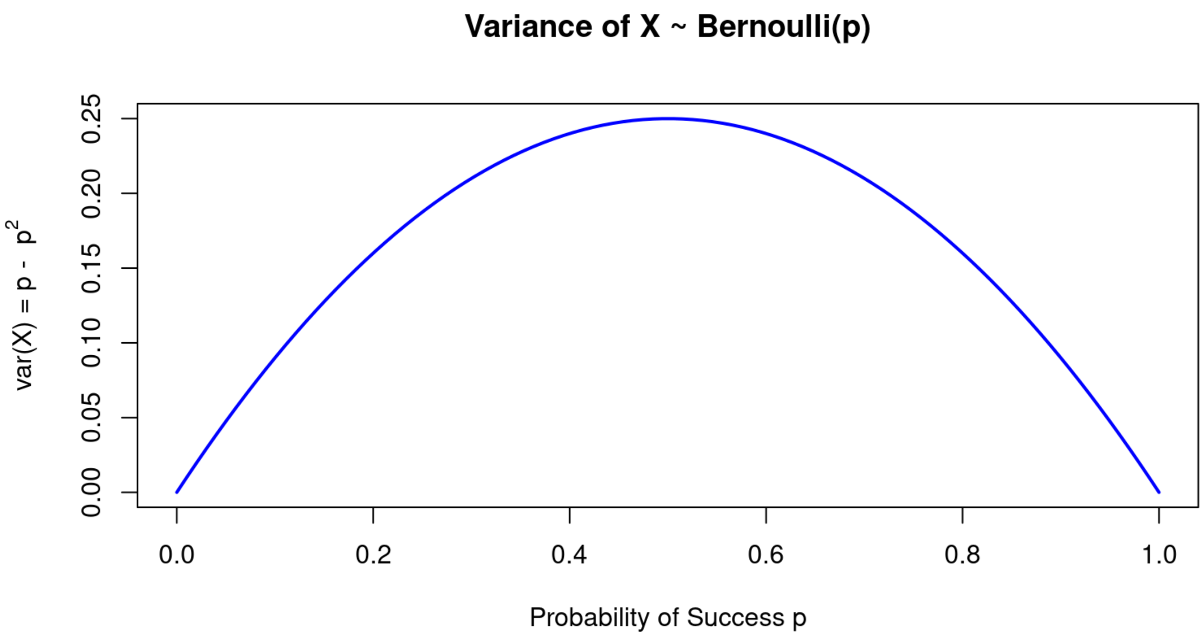

Figure 3.10

Example 3.10. Consider $X \sim \text{Bern}(p)$. We know $E(X) = p$, so by LOTUS, we know that

$$\begin{aligned}
\text{var}(X) &= E[(X - E(X))^2] \\
&= E[(X - p)^2] \\
&= \sum_{k=0}^{1}(k - p)^2 P(X = k) \\
&= (0 - p)^2(1 - p) + (1 - p)^2 p \\
&= p^2(1 - p) + (1 - p)^2 p \\
&= p(1 - p)(p + 1 - p) \\
&= p(1 - p) \\
&= p - p^2.
\end{aligned}$$

Figure 3.10 plots the variance of the Bernoulli as a function of the probability of success p. We observe a few things: if $p = 0$, meaning $X = 0$ is guaranteed, or $p = 1$, meaning $X = 1$ is guaranteed, then the variance is 0, meaning X is perfectly predictable and there are no surprises. If $p = 1/2$, then the variance is maximized. Thinking of X as a coin flip then, this implies that a fair coin is the most surprising, or least predictable, which makes sense. It could go either way with equal probability.

In the Bernoulli example, we computed the variance using LOTUS. That is the last time we will do that. The following **computation formula** is much more convenient.

Theorem 3.4. Let X be any random variable with finite mean and variance. Then

$$\operatorname{var}(X) = E(X^2) - E(X)^2. \tag{3.11}$$

Proof. This calculation is an exercise in applying the linearity of expectation, whilst remembering that, whatever its value happens to be, $E(X)$ is itself just a constant. So you should treat it like one. Observe:

$$\begin{aligned}
\operatorname{var}(X) &= E[(X - E(X))^2] \\
&= E\left[X^2 - 2E(X)X + E(X)^2\right] \\
&= E(X^2) - E[2E(X)X] + E[E(X)^2] \\
&= E(X^2) - 2E(X)E(X) + E(X)^2 \\
&= E(X^2) - 2E(X)^2 + E(X)^2 \\
&= E(X^2) - E(X)^2.
\end{aligned}$$

□

Remark 3.7. The quantity $E(X^2)$ is called the **second moment** of X. But there is nothing special about the number two, and in general $E(X^n)$ is called the n**th moment** of X.

Example 3.11. Let $X \sim \text{Bern}(p)$ again. We can use LOTUS to compute the second moment:

$$E(X^2) = \sum_{k=0}^{1} k^2 P(X = k) = 0^2(1 - p) + 1^2 p = p.$$

With this, Theorem 3.4 gives that

$$\operatorname{var}(X) = E(X^2) - E(X)^2 = p - p^2 = p(1 - p).$$

This is the same result that we got in Example 3.10, but the computation was cleaner.

Example 3.12. The second moment of $X \sim \text{Poisson}(\lambda)$ is

$$
\begin{aligned}
E(X^2) &= \sum_{k=0}^{\infty} k^2 e^{-\lambda} \frac{\lambda^k}{k!} \\
&= \sum_{k=1}^{\infty} k^2 e^{-\lambda} \frac{\lambda^k}{k!} \\
&= \lambda e^{-\lambda} \sum_{k=1}^{\infty} k^2 \frac{\lambda^{k-1}}{k!} \\
&= \lambda e^{-\lambda} \sum_{k=1}^{\infty} k \frac{\lambda^{k-1}}{(k-1)!} \\
&= \lambda e^{-\lambda} \sum_{k=1}^{\infty} (k-1+1) \frac{\lambda^{k-1}}{(k-1)!} \\
&= \lambda e^{-\lambda} \sum_{k=1}^{\infty} \left[(k-1) \frac{\lambda^{k-1}}{(k-1)!} + \frac{\lambda^{k-1}}{(k-1)!} \right] \\
&= \lambda e^{-\lambda} \left[\sum_{k=1}^{\infty} (k-1) \frac{\lambda^{k-1}}{(k-1)!} + \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} \right] \\
&= \lambda e^{-\lambda} \left[\sum_{k=2}^{\infty} (k-1) \frac{\lambda^{k-1}}{(k-1)!} + \sum_{i=0}^{\infty} \frac{\lambda^{i}}{i!} \right] \\
&= \lambda e^{-\lambda} \left[\sum_{k=2}^{\infty} \frac{\lambda^{k-1}}{(k-2)!} + e^{\lambda} \right] \\
&= \lambda e^{-\lambda} \left[\lambda \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!} + e^{\lambda} \right] \\
&= \lambda e^{-\lambda} \left[\lambda \sum_{j=0}^{\infty} \frac{\lambda^{j}}{j!} + e^{\lambda} \right] \\
&= \lambda e^{-\lambda} \left[\lambda e^{\lambda} + e^{\lambda} \right] \\
&= \lambda^2 e^{-\lambda} e^{\lambda} + \lambda e^{-\lambda} e^{\lambda} \\
&= \lambda^2 + \lambda.
\end{aligned}
$$

So $\text{var}(X) = E(X^2) - E(X)^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$, and strangely, the Poisson distribution has $E(X) = \text{var}(X)$. We saw this in Figure 3.7, where the pmf shifted rightward *and widened* as λ increased. This property renders the Poisson distribution less useful for modeling count-valued data than we might hope, because we cannot separately manipulate the center and the spread of the distribution. An entire literature has emerged in statistics that attempts to modify the Poisson so that it does not have this funky property.

When studying the expected value, we wondered how we might compute the expectation of a transformation $E[g(X)]$, and we came up with LOTUS. Is there a convenient, LOTUS-like formula for var$[g(X)]$? In general, no. But an important special case is the linear transformation $g(X) = aX + b$. We know that the expected value is a linear operator, and so $E(aX + b) = aE(X) + b$. There is a clean formula for var$(aX + b)$, but it's *not* linear:

Theorem 3.5. Let X be any random variance with finite mean and variance, and let $a, b \in \mathbb{R}$ be arbitrary constants. Then

$$\operatorname{var}(aX + b) = a^2 \operatorname{var}(X). \tag{3.12}$$

Proof.

$$\begin{aligned}
\operatorname{var}(aX + b) &= E[(aX + b)^2] - E(aX + b)^2 \\
&= E(a^2X^2 + 2abX + b^2) - [aE(X) + b]^2 \\
&= E(a^2X^2 + 2abX + b^2) - [a^2E(X)^2 + 2abE(X) + b^2] \\
&= a^2E(X^2) + 2abE(X) + b^2 - a^2E(X^2) - 2abE(X) - b^2 \\
&= a^2E(X^2) - a^2E(X)^2 \\
&= a^2[E(X^2) - E(X)^2] \\
&= a^2\operatorname{var}(X).
\end{aligned}$$

□

Remark 3.8. As we see, it is not the case that var$(aX + b)$ is equal to $a$var$(X) + b$, and so the variance *is not* a linear operator. But this makes sense. The variance is a measure of spread, and merely shifting the location of a random variable with $X + b$, for instance, should not have an effect on how spread out it is. The spread remains the same.

Theorem 3.6. Let $X_1, X_2, ..., X_n$ be *independent* random variables each with finite mean and variance (possibly all different), and let $a_1, a_2, ..., a_n \in \mathbb{R}$ be arbitrary constants. Then

$$\operatorname{var}\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i^2 \operatorname{var}(X_i). \tag{3.13}$$

Example 3.13. If $I_i \overset{\text{iid}}{\sim} \text{Bern}(p)$, then $X = \sum_{i=1}^n I_i \sim \text{Binom}(n, p)$, and we used this fact together with the linearity of expectation to show that $E(X) = np$. We can now use Example 3.10 and Theorem 3.6 to perform a quick derivation of the variance of a binomial random variable:

$$\operatorname{var}(X) = \operatorname{var}\left(\sum_{i=1}^{n} I_i\right) = \sum_{i=1}^{n} \operatorname{var}(I_i) = \sum_{i=1}^{n} p(1-p) = np(1-p).$$