Chapter 7.2

Variance of a Discrete Random Variable

Review of Last Week

Binomial random variable: counts the number of successes in n **independent** trials with **constant** probability of success equal to π on each trial.

$$X \sim Binom(n, \pi)$$

$$f(x) = \binom{n}{x} \pi^x (1-\pi)^{n-x}, \quad x = 0, 1, 2, \ldots, n$$

Binomial calculations in R:

- `dbinom(x, size, prob)` calculates $P(X = x)$
- `pbinom(q, size, prob)` calculates $P(X \leq q)$
- `pbinom(q, size, prob, lower.tail = F)` calculates $P(X > q)$

Review of Last Week

Expected Value (denoted by μ or $E[X]$) average value of the random variable across many replications of the experiment.

- $E[X] = \sum_x x \cdot f(x)$
- $E[X] = n\pi$ when $X \sim Binom(n, \pi)$.
- Law of the unconscious probabilist: In general for any $Y = t(X)$

$$E[Y] = \sum_x t(x) f(x).$$

- As a special case, when $Y = aX + b$ we get the result

$$E[Y] = aE[X] + b.$$

We refer to this result as **linearity of expected value**

Example: X = daily demand for a certain product
The PMF of X is shown below

x	1	5	10
$f(x)$	8/10	1/10	1/10

The expected daily demand is

$\mu = E[X] = 1 \times \frac{8}{10} + 5 \times \frac{1}{10} + 10 \times \frac{1}{10} = 2.3$ units ←

Now define Y = daily profit from selling the product
$= (1 - e^{-2X})$ (non-linear fn of X)

Find the expected daily profit
Brute force method: Find the PMF of Y, then use it to find E(Y)

y	$(1-e^{-2\times1})$	$(1-e^{-2\times5})$	$(1-e^{-2\times10})$	→ $1-e^{-2x}$
$f_Y(y)$	8/10	1/10	1/10	→ $f(x)$

$E(Y) = (1-e^{-2\times1}) \times \frac{8}{10} + (1-e^{-2\times5}) \times \frac{1}{10} + (1-e^{-2\times10}) \cdot \frac{1}{10} = \$.89$

$= \sum_x \underbrace{(1-e^{-2x})}_{t(x)} f(x)$

In general for $Y = t(X)$ Law of Unconscious Probabilist
$E(Y) = \sum_x t(x) f(x)$

Special case $Y = t(X) = aX + b$ (linear fn)

$E(Y) = \sum_x (ax+b) f(x) = \sum_x (a.x.f(x) + b.f(x))$

$= \sum_x a.x.f(x) + \sum_x b.f(x)$

$= a \sum_x x.f(x) + b \underbrace{\cancel{\sum_x f(x)}}_{1} = aE(X) + b$

Variation
$Y = \ln(X)$
What is $E(Y)$?

$E(Y) = \sum_x \ln(x).f(x)$

$= \ln(1).\frac{8}{10} + \ln(5).\frac{1}{10} + \ln(10).\frac{1}{10}$

$Y = aX^2 + bX + c$

$E(Y) = \sum_x (ax^2 + bx + c) f(x)$

$= \sum_x ax^2 f(x) + \sum_x bx f(x) + \sum_x c f(x)$

$= a E(X^2) + b E(X) + c$

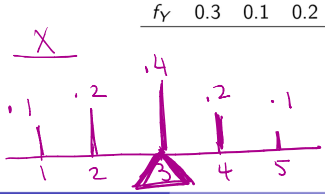
Warm up

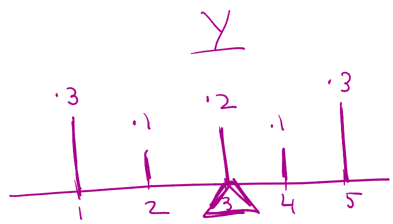
$$E(Y) = \sum_x (ax+b)\, f(x) = \sum_x \ldots$$
$$= \sum_x a.x.f(x) + \sum_x b.f(x)$$
$$= a \sum_x x.f(x) + b \cancel{\sum_x f(x)}^{\,1} = aE(X)+b$$

Warm up

Let's consider two random variables X and Y with PMF as follows. Both have an expected value of 3. But which will deviate more from expected?

x	1	2	3	4	5
f_X	0.1	0.2	0.4	0.2	0.1
f_Y	0.3	0.1	0.2	0.1	0.3





$$E(X) = \sum x.f_X(x) = 1\times 0.1 + 2\times 0.2 + 3\times 0.4 + 4\times 0.2 + 5\times 0.1 = \boxed{3}$$

$$E(Y) = \boxed{3}$$

Y deviates from its mean more than X will.

Variance of a random variable

Definition 7.2 Let X be a discrete random variable with mean μ. The **variance** of X, assuming it exists, is defined by

$$\sigma_X^2 = \underset{V(X)}{Var[X]} = E\left[\overbrace{(X-\mu)^2}^{t(X)}\right] = \sum_x (x-\mu)^2 \times f(x).$$

The **standard deviation** is the positive square root of the variance.

$X - \mu$: deviation of X from its mean

- The interpretation attached to the variance is that small values imply that X is very likely to be close to $E[X]$ and larger values imply X is more variable.
- The standard deviation has this same qualitative interpretation but is easier to interpret in that the unit of measurement is the same as that for the original random variable. (The unit of measurement on the variance is the square of the original unit.)

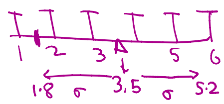
An alternate formula for $Var[X]$ that is not as intuitive as the definition, but is often easier to use in practice.

$$
\begin{aligned}
Var[X] &= E\left[(X-\mu)^2\right], \quad \longrightarrow \text{ original defn for variance} \\
&= \sum_x (x-\mu)^2 \cdot f(x), \\
&= \sum_x (x^2 - 2x\mu + \mu^2) \cdot f(x), \\
&= \sum_x x^2 \cdot f(x) - \sum_x 2x\mu f(x) + \sum_x \mu^2 f(x) \\
&= E\left[X^2\right] - 2\mu \sum_x x f(x) + \mu^2 \sum_x f(x), \\
&= E\left[X^2\right] - 2\mu \cdot \mu + \mu^2 \cdot 1, \\
&= \boxed{E\left[X^2\right] - \mu^2.} \quad \longrightarrow \text{ shortcut way of calculation}
\end{aligned}
$$

Example 7.6

Calculate *Var* [*X*] if *X* represents the outcome when a fair die is rolled.

$Var(X) = E(X^2) - \mu^2$



Find the PMF:

x	1	2	3	4	5	6
$f(x)$	1/6	1/6	1/6	1/6	1/6	1/6

From last time: $\mu = E(X) = 3.5$

Lets find $E(X^2) = \sum_x x^2 \cdot f(x)$ (by law of un. prob)

$= 1^2 \cdot \frac{1}{6} + 2^2 \cdot \frac{1}{6} + 3^2 \cdot \frac{1}{6} + 4^2 \cdot \frac{1}{6} + 5^2 \cdot \frac{1}{6} + 6^2 \cdot \frac{1}{6}$

$= \boxed{\frac{91}{6}}$

$\sigma^2 = Var(X) = \frac{91}{6} - 3.5^2 = \boxed{2.917}$

$\sigma = SD(X) = +\sqrt{Var(X)} = \sqrt{2.917} = \boxed{1.708}$

So X will typically vary by 1.708 from its expected value of 3.5

Example 7.7

$f(x) = \binom{n}{x} \pi^x (1-\pi)^{n-x}, \; x = 0, , n$

If $X \sim Binom(1, \pi)$, then X is called a **Bernoulli** random variable. For a Bernoulli random variable, what are $E[X]$ and $Var[X]$?

x	0	1
$f(x)$	$1-\pi$	π

$\mu = E(X) = \sum_x x \cdot f(x) = 0 \cdot (1-\pi) + 1 \cdot \pi = \pi$

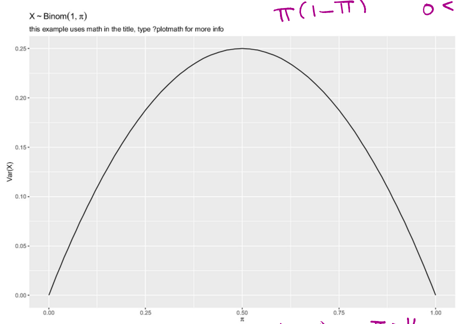
$\sigma^2 = Var(X) = E(X^2) - \mu^2$

$E(X^2) = \sum_x x^2 \cdot f(x) = 0^2 \cdot (1-\pi) + 1^2 \cdot \pi = \pi$

$\sigma^2 = \pi - \pi^2 = \pi(1-\pi)$

As a function of π, $Var(X)$ in example 7.7 is quadratic. Its graph is a parabola that opens downward and the largest variance occurs when $\pi = \frac{1}{2}$.

$\pi(1-\pi)$ $0 < \pi < 1$



$\pi(1-\pi)$ achieves its max value when $\pi = 1/2$

$X \sim Binom(n, \pi)$

$Var(X) = n\pi(1-\pi)$

(This is relevant for HW 4 problem 3)

```
ggplot() +
  geom_function(fun = function(x){x*(1-x)}, xlim = c(0,1) ) +
  labs(x = expression(pi),
       y = "Var(X)",
       title = expression(X %~% Binom(1, pi)),
       subtitle = "type ?plotmath for more info" )
```

Variance of a Linear Transformation

Lemma 7.4 Let X be a discrete random variable and suppose a and b are constants. Then

$$Var(aX + b) = a^2 Var(X).$$

Example 7.8

Suppose $E(X) = \mu$ and the standard deviation of X equals σ. Find the mean and variance of the new random variable

$$Y = (X - \mu)/\sigma.$$

Z score

$$Y = \underbrace{\tfrac{1}{\sigma}}_{a} \cdot X + \underbrace{\left(\tfrac{-\mu}{\sigma}\right)}_{b}$$

$$E(Y) = E\left(\tfrac{1}{\sigma} X + \left(\tfrac{-\mu}{\sigma}\right)\right) = \tfrac{1}{\sigma} \cancel{E(X)}^{\mu} - \tfrac{\mu}{\sigma} = 0$$

Linearity of expectation
$E(aX+b) = aE(X)+b$

$$Var(Y) = Var\left(\tfrac{1}{\sigma} X + \left(\tfrac{-\mu}{\sigma}\right)\right) = \tfrac{1}{\sigma^2} Var(X)$$

$$= \tfrac{1}{\sigma^2} \times \sigma^2 = 1$$

$Var(aX+b) = a^2 Var(X)$
$SD(aX+b) = \sqrt{Var(aX+b)}$
$= \sqrt{a^2 Var(X)}$
$= |a| SD(X)$

$SD(Y) = \sqrt{Var(Y)} = 1$

The fact that

$$Y = (X - \mu)/\sigma$$

has mean 0 and standard deviation 1 is the reason it is often referred to as the *standardization* of X.

Chebyshev's Inequality

Lemma 7.5 For any random variable X with mean μ and standard deviation σ:

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2},$$

where k is some positive number.

Note: $|X - \mu| \geq k\sigma$ means either

$$(X - \mu) \geq k\sigma \Rightarrow X \geq \mu + k\sigma$$

or

$$(X - \mu) \leq -k\sigma \Rightarrow X \leq \mu - k\sigma.$$

That is, Chebyshev's inequality guarantees that at most $\frac{1}{k^2}$ of the distribution can be k or more standard deviations from the mean

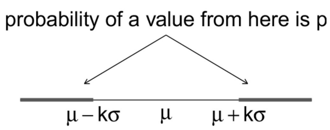
Tshebysheff

Chebyshev's Inequality: heuristic proof

Let us denote

$$p = P(|X - \mu| \geq k\sigma). \quad (1)$$

For any X which satisfies equation (1)



Chebyshev (heuristic proof)

One PMF which satisfies equation (1) is the following:

value	$\mu - k\sigma$	μ	$\mu + k\sigma$
probability	$\frac{p}{2}$	$1 - p$	$\frac{p}{2}$

This PMF yields the smallest variance that is possible for any distribution that satisfies 1.

Chebyshev's Inequality: heuristic proof

It is straightforward to check that the variance of this distribution

value	$\mu - k\sigma$	μ	$\mu + k\sigma$
probability	$\frac{p}{2}$	$1-p$	$\frac{p}{2}$

is $pk^2\sigma^2$ and therefore

$$Var(X) = \sigma^2 \geq pk^2\sigma^2$$

from which it follows that

$$p = P(|X-\mu| \geq k\sigma) \leq \frac{1}{k^2}.$$

Example 7.9

What does Chebyshev's inequality say about the probability that a random variable is more than 2 standard deviations from its mean? Calculate the exact probability for the outcome, X, when a fair die is rolled.

Say $E(X) = \mu$ and $Var(X) = \sigma^2$, $k = 2$

Chebyshev's inequality says that

$P(|X-\mu| \geqslant 2\sigma) \leqslant \frac{1}{2^2}$

Ex: X is a number on a fair die

From earlier: $\mu = E(X) = 3.5$, $\sigma = SD(X) = 1.7$

$P(|X-3.5| \geqslant 2 \times 1.7) \leqslant \frac{1}{4}$

$\Rightarrow P(X - 3.5 \geqslant 3.4 \cup X - 3.5 \leqslant -3.4) \leqslant \frac{1}{4}$

$|x| \geqslant 3$

$x \geqslant 3$

$x \leqslant -3$

$\Rightarrow P(X \geqslant 6.9 \cup X \leqslant 0.1) \leqslant 1/4$

Exact value is 0 for this probability when X is # on a die roll.

Variation: $k = 1.1$

Chebyshev's inequality says

$P(|X-\mu| \geqslant 1.1\sigma) \leqslant \frac{1}{1.1^2} = 0.826$

Chebyshev's inequality says
$P(|X-\mu| \geq 1.1\sigma) \leq \frac{1}{1.1^2} = 0.826$

X: # on a die roll
$P(|X-3.5| \geq \underbrace{1.1 \times 1.7}_{1.87}) = P((X-3.5) \geq 1.87 \cup (X-3.5) \leq -1.87)$
$= P(X \geq 5.37 \cup X \leq 1.63)$
$= P(X = 6 \cup X = 1) = \boxed{\frac{1}{3}}$

Variation 2: Random variable X has
$\mu = 5, \quad \sigma = 2$
what can you say about $P(-1 < X < 11)$

$P(-1 < X < 11) = P(-1-5 < X-5 < 11-5)$
$= P(-6 < (X-5) < 6)$
$= P(|X-5| < 6) = 1 - P(|X-5| \geq 6)$ ↳ 3×2 $(k=3)$
$\geq 1 - \frac{1}{3^2} = \frac{8}{9}$

Chebyshev's inequality says
$P(|X-5| \geq k.2) \leq \frac{1}{k^2}$
$-P(|X-5| \geq 2k) \geq -\frac{1}{k^2}$

Chebyshev's inequality only provides informative answers for values of k greater than 1. In addition, the bound $\frac{1}{k^2}$ is usually much larger than the exact probability for many distributions.