Topic 2.3: Moments

We do 4 things in this class

① Build models

* counting when all elements of discrete sample space are equally likely
* conditional events
* any $g(x) \geq 0$ that is piecewise continuous with finite integral.

② Compute probabilities for a given model

* Axioms of probability
* pmf, pdf, cdf

③ Learn

* Bayes Rule

④ compute summary statistics

- expected value, variance, moments

Topic 2.3: Moments of RVs (Chapter 3.3 and 4.3)

Once we know $p_X(x)$, $f_X(x)$, or $F_X(x)$, we know <u>everything</u> about the probability model for the RV X.

But it may take a lot of detail to convey.

Sometimes, we just care about summary information.

minimum
maximum
* mean
mode — most common value
median — "middle": half of experiments will have an outcome larger than the median
percentile (quartile)
* standard deviation
* variance
* n^{th} moments

Expected value (mean)

$$m_X = \mu_X = E(X)$$

a real number, even if X is discrete

$$E(X) = \sum_{x \in S_X} x \, p_X(x)$$ for discrete X, when $S_X = \{x_1, x_2, \dots\}$

$$E(X) = \int_{-\infty}^{\infty} x \, f_X(x) \, dx$$ for continuous X

(Defined only if this sum or integral converges absolutely)

Examples (discrete RV)

Bernoulli RV $\quad p_X(x) = \begin{cases} 1-p & x=0 \\ p & x=1 \end{cases}$

$$E(X) = \sum_{x \in S_X} x \, p_X(x)$$

$$= 0 \cdot (1-p) + 1 \cdot p = p$$

Uniform Discrete RV

$p_X(x) = \frac{1}{M}$ for each $x \in S_X$

Suppose $S_X = \{0, 1, \ldots, M-1\}$

$$E(X) = \sum_{x \in S_X} x \, p_X(x) = 0 \cdot \frac{1}{M} + 1 \cdot \frac{1}{M} + \cdots + (M-1)\frac{1}{M}$$

$$= \frac{1}{M} \sum_{i=0}^{M-1} i = \frac{1}{M} \frac{M(M-1)}{2} = \frac{M-1}{2}$$

Note: if M is even, say, $M=6$, then

$E(X) = \frac{6-1}{2} = \frac{5}{2}$. This is OK, even though $E(X)$ is not a member of S_X.

Examples (discrete RV)

The 3-game series, alternating home and away.

X = # times team C wins

Y = # games the teams play

$$p_X(x) = \begin{cases} p(1-p) & x=0 \\ p^2(1-p) + (1-p)^3 & x=1 \\ p(1-p) + p^3 + (1-p)^2 p & x=2 \end{cases}$$

$$p_Y(y) = \begin{cases} 2p(1-p) & y=2 \\ 1-2p(1-p) & y=3 \end{cases}$$

$$E(X) = \sum_{x \in S_X} x \, p_X(x)$$

$$= 0 \cdot p(1-p) + 1 \cdot [p^2(1-p) + (1-p)^3] + 2[p(1-p) + p^3 + (1-p)^2 p]$$

$$= p^2 - p^3 + 1 - 3p + 3p^2 - p^3 + 2p - 2p^2 + 2p^3 + 2p - 4p^2 + 2p^3$$

$$= 2p^3 - 2p^2 + p + 1 \quad \text{(if I did the math right)}$$

$$E(Y) = \sum_{y \in S_Y} y \, p_Y(y) = 2[2p(1-p)] + 3[1-2p(1-p)]$$

$$= 4p(1-p) - 6p(1-p) + 3 = 3 - 2p(1-p)$$

Examples (continuous RV)

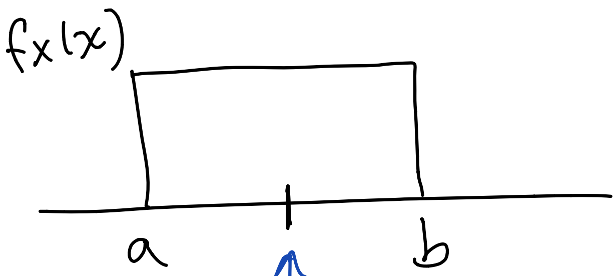
Continuous uniform RV (section 4.4.1)

A uniform RV in the interval $[a,b]$ has PDF

$$f_X(x) = \begin{cases} 1/(b-a) & a \le x \le b \\ 0 & \text{else} \end{cases}$$

$$E(X) = \int_{-\infty}^{\infty} x \, f_X(x)\, dx = \int_a^b \frac{x}{b-a}\, dx$$

$$= \left.\frac{x^2}{2(b-a)}\right|_a^b = \frac{b^2-a^2}{2(b-a)} = \frac{b+a}{2} \quad \text{if } b \ne a.$$



$E(X) = \frac{b+a}{2}$ the middle.

Intuitively, $E(X)$ is the balance point, the center of mass.

Examples (continuous RV)

Exponential RV

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & \text{else} \end{cases}$$

$$E(X) = \int_{-\infty}^{\infty} x\, f_X(x)\, dx = \int_0^{\infty} \lambda x e^{-\lambda x}\, dx$$

Integrate by parts. $\int u\,dv = uv - \int v\,du$

$u = x \qquad dv = \lambda e^{-\lambda x} dx$

$du = dx \qquad v = -e^{-\lambda x}$

$$E(X) = -x e^{-\lambda x}\Big|_0^{\infty} + \int_0^{\infty} e^{-\lambda x}\, dx$$

$$= -x e^{-\lambda x}\Big|_0^{\infty} - \frac{1}{\lambda} e^{-\lambda x}\Big|_0^{\infty}$$

$$= -(0-0) - \frac{1}{\lambda}(0-1) = \boxed{\frac{1}{\lambda}}$$

mean of an exponential RV with parameter λ is $1/\lambda$.

Expectation of a function of a RV

(Chapter 3.3.1)

What if we know $p_X(x)$ and we have another RV $Y = g(X)$. What's $E(Y)$?

Example: X is a voltage, among the set

$$S_X = \{-2, -1, 0, 1, 2, 3, 4, 5\}$$

where all are equally likely: $p_X(x) = \begin{cases} \frac{1}{8} & x \in S_X \\ 0 & \text{else.} \end{cases}$

We know how to compute

$$E(X) = \sum_{x \in S_X} x\, p_X(x) = \frac{1}{8}\left[-2 + -1 + 0 + 1 + 2 + 3 + 4 + 5\right] = \frac{12}{8} = \frac{3}{2}$$

Let $Y = X^2$. Find $E(Y) = m_Y$.

Method 1: Find $S_Y = \{0, 1, 4, 9, 16, 25\}$

① Find $p_Y(y)$

x	y	$p_Y(y)$
0	0	1/8
−1, +1	1	2/8
−2, +2	4	2/8
3	9	1/8
4	16	1/8
5	25	1/8

② $E(Y) = \sum_{y \in S_Y} y\, p_Y(y)$

using the table

$$= 0\left(\tfrac{1}{8}\right) + 1\left(\tfrac{2}{8}\right) + 4\left(\tfrac{2}{8}\right) + 9\left(\tfrac{1}{8}\right) + 16\left(\tfrac{1}{8}\right) + 25\left(\tfrac{1}{8}\right)$$

$$= 7.5$$

Method 2: what if you just want $E(Y)$, and don't care about $p_Y(y)$?

$$E(Y) = E(X^2) = \sum_{x \in S_X} x^2 p_X(x) = \frac{1}{8}\left[4+1+0+1+4+9+16+25\right]$$
$$= 7.5$$

More generally:

$$E(Y) = E(g(X)) = \sum_{x \in S_X} g(x) p_X(x)$$ discrete RVs

$$E(Y) = E(g(X)) = \int_{-\infty}^{\infty} g(x) f_X(x)\,dx$$ continuous RVs

Why? if we group terms of x that map to the same $y = g(x)$

$$E(Y) = \sum_{y \in S_Y} y \left[\sum p_X(x)\right]$$

sum is over those x such that $g(x) = y$ (which is $p_Y(y)$)

$$= \sum_{y \in S_Y} y\, p_Y(y)$$

Law of the Unconscious Statistician (LOTUS)

Another way to look at the same example

$x^2 p_X(x)$	$p_X(x)$	x	$y=x^2$	$p_Y(y)$	$y\,p_Y(y)$
0	1/8	0	0	1/8	0
1/8	1/8	−1	1	2/8	2/8
1/8	1/8	+1			
4/8	1/8	−2	4	2/8	8/8
4/8	1/8	+2			
9/8	1/8	+3	9	1/8	9/8
16/8	1/8	+4	16	1/8	16/8
25/8	1/8	+5	25	1/8	25/8

sum 60/8 $= \sum_{x \in S_X} x^2 p_X(x)$

sum 60/8 $= \sum_{y \in S_Y} y\, p_Y(y)$

Properties of $E(x)$

Linearity : $E[g(X) + h(X)] = E[g(X)] + E(h(X))$

Scale : $E(cX) = cE(X)$ (where c is constant)

DC shift : $E(X+d) = E(X) + d$ (where d is a constant)

$E(d) = d$ (where d is a constant)

Applying properties of Expectation

Example : Let X be a noise voltage, uniformly distributed

$S_X = \{-3, -1, 1, 3\}$

Let $Y = 2X + 10$. Let $Z = Y^2$. What is $E(Z)$?

Solution : $E(Z) = E(Y^2) = E\left((2X+10)^2 \right)$

$= E(4X^2 + 40X + 100)$

$= 4E(X^2) + 40E(X) + 100$

Compute $E(X) = 0$; $E(X^2) = \frac{1}{4}(9+1+1+9) = 5$

So $E(Z) = 20 + 100 = 120$

Example: X = temperature in °F

Y = temperature in °C

$$Y = (X - 32)\frac{5}{9}$$

$$E(Y) = \frac{5}{9}\left(E(X) - 32\right)$$

Is $E(g(X)) = g(E(X))$?

In general, no! Example, $E(X^2) \neq (E(X))^2$ in general

Variance (and standard deviation) of a RV

(Chapter 3.3.2)

mean is often called the first moment of X

$E(X) = 1^{st}$ moment

$E(X^2) = 2nd$ moment

$E(X^n) = n^{th}$ moment

Variance is the 2nd central moment

$$\sigma_X^2 = VAR(X) = E((X - m_X)^2)$$

Central moments measure moments of the RV with the mean removed $\Rightarrow$ "centralized"

$\Rightarrow$ Variance describes how much X varies about its mean during different experiments.

Since $(X - m_X)^2$ is a function of X,

we can compute

$$VAR(X) = \sum_{x \in S_X} (x - m_X)^2 p_X(x)$$

$\boxed{\text{discrete RVs}}$

$$VAR(X) = \int_{-\infty}^{\infty} (x - m_X)^2 f_X(x)\, dx$$

$\boxed{\text{continuous RVs}}$

Standard deviation

$$\sigma_X = STD(X) = \sqrt{VAR(X)}$$

why standard deviation and not variance?

Units! σ_X^2 has units (units of X)^2

σ_X has units (units of X)

unit: meters, feet, pounds, Kg/m^2

Example: the temperature in July is on average 85°F with a standard deviation of 10°F

A short cut for computing variance

$$\begin{aligned} Var(X) &= E((X-m_X)^2) \\ &= E(X^2 - 2Xm_X + m_X^2) \\ &= E(X^2) - 2m_X E(X) + m_X^2 \end{aligned}$$

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

I'll often express this as

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

$$\boxed{\text{Note: } E(X^2) = Var(X) + E(X)^2 = \sigma^2 + \mu^2}$$

Warning about computing variance on a computer

method 1

```
sum1 = 0
sum2 = 0
for i = 1:n
    sum1 += x(n)
    sum2 += x(n)^2
end
sum1 /= n
sum2 /= n
var = sum2 - sum1*sum1
```

Implements

$$E(X^2) - E(X)^2$$

method 2

```
sum1 = 0
for i = 1:n
    sum1 += x(n)
end
sum1 /= n

sum2 = 0
for i = 1:n
    sum2 += (x(n) - sum1)^2
end
sum2 /= n

var = sum2
```

Implements

$$E\left(\left(X - E(X)\right)^2\right)$$

If n is very large, why might these not give the same numerical result? Which one will be more accurate?

(The difference is also magnified when most x(n) are near E(X))

Variance examples

Bernoulli RV. Recall $X_A = \begin{cases} 1 & \text{event } A \text{ occurs} \\ 0 & \text{else} \end{cases}$

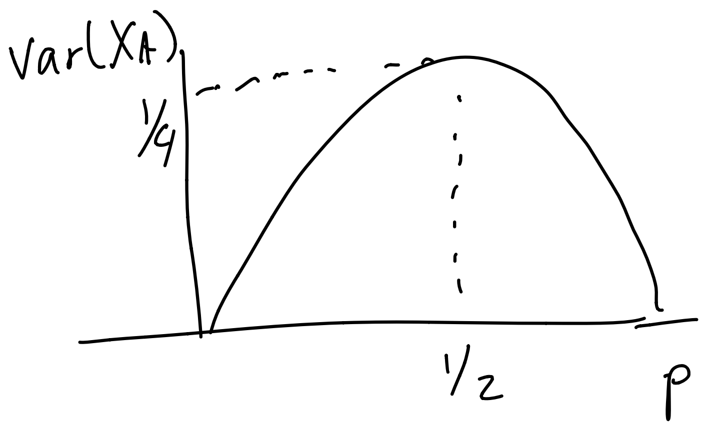
$S_X = \{0, 1\}$

$p_X(1) = P(A) = p$ $\qquad$ $p_X(0) = 1 - p$

$E(X_A) = 0 \cdot p_X(0) + 1 \cdot p_X(1) = p.$

$$\begin{aligned} Var(X_A) &= E(X^2) - E(X)^2 \\ &= \left[0^2 p_X(0) + 1^2 p_X(1) \right] - p^2 \\ &= p - p^2 = \boxed{p(1-p)} \end{aligned}$$

Coin toss w/ a biased coin



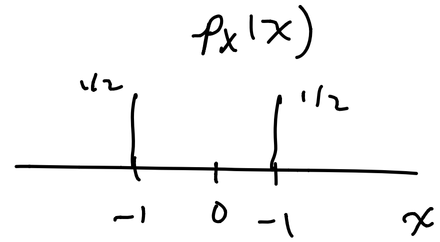
(symmetric)

Variance of a Bernoulli RV is small if p is small or if p is near 1.

Variance is largest when $p = 1/2$

Variance examples

a) $$p_X(x) = \begin{cases} 1/2 & x = +1 \\ 1/2 & x = -1 \\ 0 & \text{else} \end{cases}$$

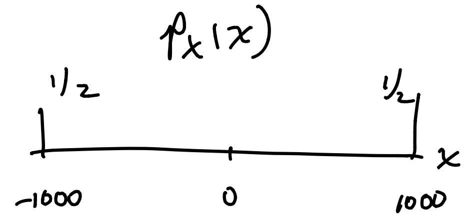


$m_X = 0$

$$Var(X) = E(X^2) - m_X^2 = E(X^2)$$

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

b) $$p_X(x) = \begin{cases} 1/2 & x = +1000 \\ 1/2 & x = -1000 \\ 0 & \text{else} \end{cases}$$



$m_X = 0$

$$Var(X) = E(X^2) - m_X^2 = E(X^2)$$

$$= \frac{1}{2}(+1000)^2 + \frac{1}{2}(-1000)^2$$

$$= 1{,}000{,}000$$

$$STD(X) = 1000$$

c) $X =$ # heads in 3 tosses of a fair coin

$$E(X) = \sum_{i=0}^{3} i \binom{3}{i} \left(\frac{1}{2}\right)^i \left(\frac{1}{2}\right)^{3-i}$$

$$= 0\left(\frac{1}{8}\right) + 1\left(\frac{3}{8}\right) + 2\left(\frac{3}{8}\right) + 3\left(\frac{1}{8}\right) = \frac{12}{8} = 1.5$$

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

$$= \left[0^2\left(\frac{1}{8}\right) + 1^2\left(\frac{3}{8}\right) + 2^2\left(\frac{3}{8}\right) + 3^2\left(\frac{1}{8}\right)\right] - \left(\frac{3}{2}\right)^2$$

$$= \frac{24}{8} - \frac{9}{4} = \frac{6}{8} = \frac{3}{4}$$

Properties of Variances

① $Var(X) \geq 0$ always

② $Var(X+c) = Var(X)$ if c is constant

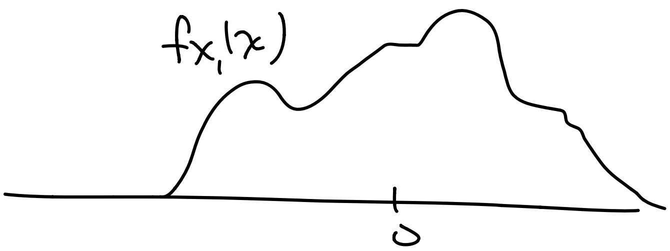
proof: $Var(X+c) = E((X+c)^2) - E(X+c)^2$

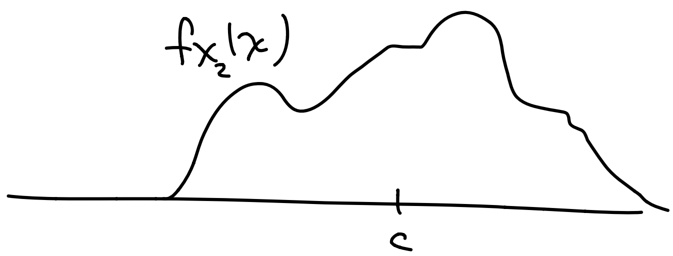
$= E(X^2+2cX+c^2) - [E(X)^2 + 2cE(X) + c^2]$

$= E(X^2) + 2cE(X) + c^2 - E(X)^2 - 2cE(X) - c^2$

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

This makes sense. Adding a constant doesn't change how much variation there is about the mean





③ $Var(cX) = c^2 Var(X)$ if c is a constant

proof: $Var(cX) = E((cX)^2) - E(cX)^2$

$= E(c^2X^2) - c^2E(X)^2 = c^2[E(X^2) - E(X)^2]$

$= c^2 Var(X)$

④ $Var(c) = 0$ if c is a constant

Comparison of properties

mean	variance
$E(aX) = aE(X)$	$Var(aX) = a^2 Var(X)$
$E(X+Y) = E(X) + E(Y)$	$Var(X+Y) \neq Var(X) + Var(Y)$ (more on this later)
$E(X+a) = E(X) + a$ if a is constant	$Var(X+a) = Var(X)$ if a is constant
$X_{min} \leq E(X) \leq X_{max}$	$Var(X) \geq 0$