Chapter 4 Expectations of functions of RVs

Example: Toss a die and win X^2 if X is # of spots.

$$E(X^2) = \sum_x x^2 p_X(x) = 1^2 \cdot \frac{1}{6} + 2^2 \cdot \frac{1}{6} + \dots + 6^2 \cdot \frac{1}{6}$$

$$= 15\tfrac{1}{6}$$

Note $E(X^2) \neq \left(E(X)\right)^2$

$$\left(E(X)\right)^2 = \left(1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + \dots + 6 \cdot \frac{1}{6}\right)^2$$

$$= (3.5)^2 = 12\tfrac{1}{4}$$

In general: $E(g(X)) = \sum_{x} g(x)\, p_X(x)$

or

$= \int g(x)\, f_X(x)\, dx$

Works for function of random vectors:

Let $Y = g(X_1, X_2, \ldots, X_k)$

with pmf $p(x_1, x_2, \ldots, x_k)$

then $E(Y) = E(g(X_1, \ldots, X_k)) = \sum_{x's} g(x_1, \ldots, x_k)\, p(x_1, \ldots, x_k)$

or $E(Y) = \int \left[\int \cdots \int g(x_1, \ldots, x_k) f(x_1, \ldots, x_k)\, dx_k \cdots dx_2 \right] dx_1$

if the integral (sum) with $|g|$ converges.

p.124 Corollary A

If X & Y are independent, then

$$E(g(X)h(Y)) = E(g(X)) \times E(h(Y))$$

provided ↑ and ↑ exist.

Special case: If X & Y are independent

$$E(XY) = E(X)E(Y)$$

if $\uparrow$ and $\uparrow$ exist.

Beware: 1) not true in general

2) converse not true

<u>Linear combinations of R.V.'s</u>

$$Y = a + b_1 X_1 + b_2 X_2 + b_3 X_3 + \cdots + b_n X_n$$

then $E(Y) = a + b_1 E(X_1) + b_2 E(X_2) + \cdots + b_n E(X_n)$

if $\uparrow$ $\cdots$ exist.

Example A (p. 126)

What is $E(Y)$ if $Y \sim \text{Bin}(n, p)$?

$$E(Y) = \sum_y \binom{n}{y} y \, p^y (1-p)^{n-y}$$

Easier: Use fact that $Y = X_1 + \cdots + X_n$

where X_i's are indep. $\text{Bin}(1, p)$

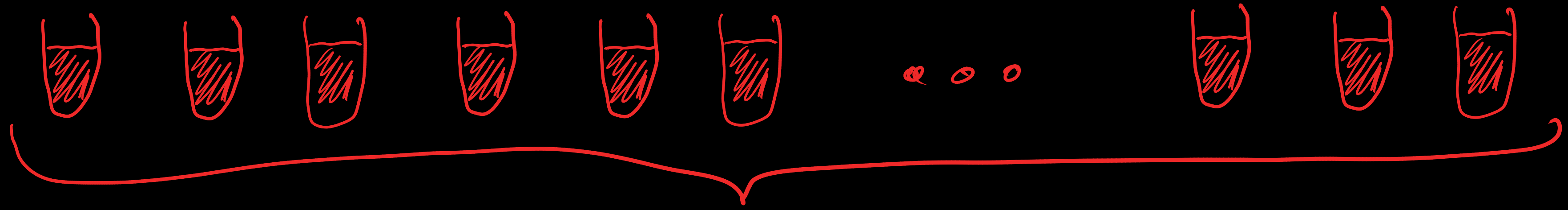
and $E(X_i) = p$

So $E(Y) = \sum_{i=1}^{n} E(X_i) = np$

Example C p.128

Group testing

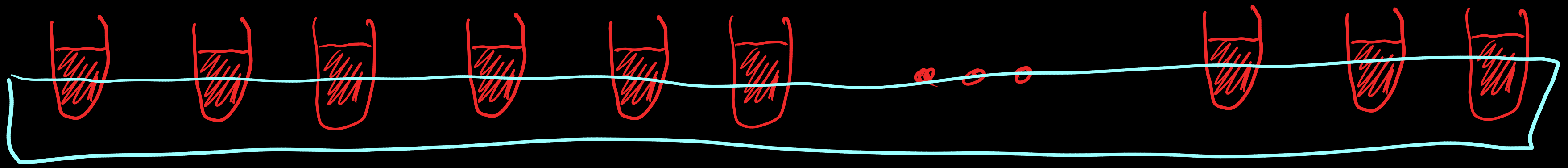
- n blood samples tested for rare disease



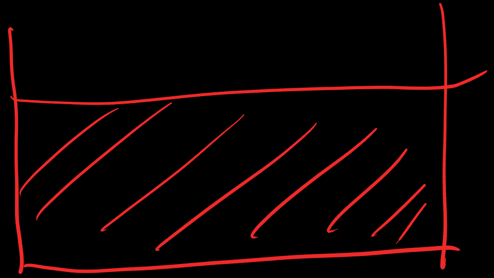
n

- If you test each individually will need n tests.

Alternative:



take $1/2$ of each sample and combine:



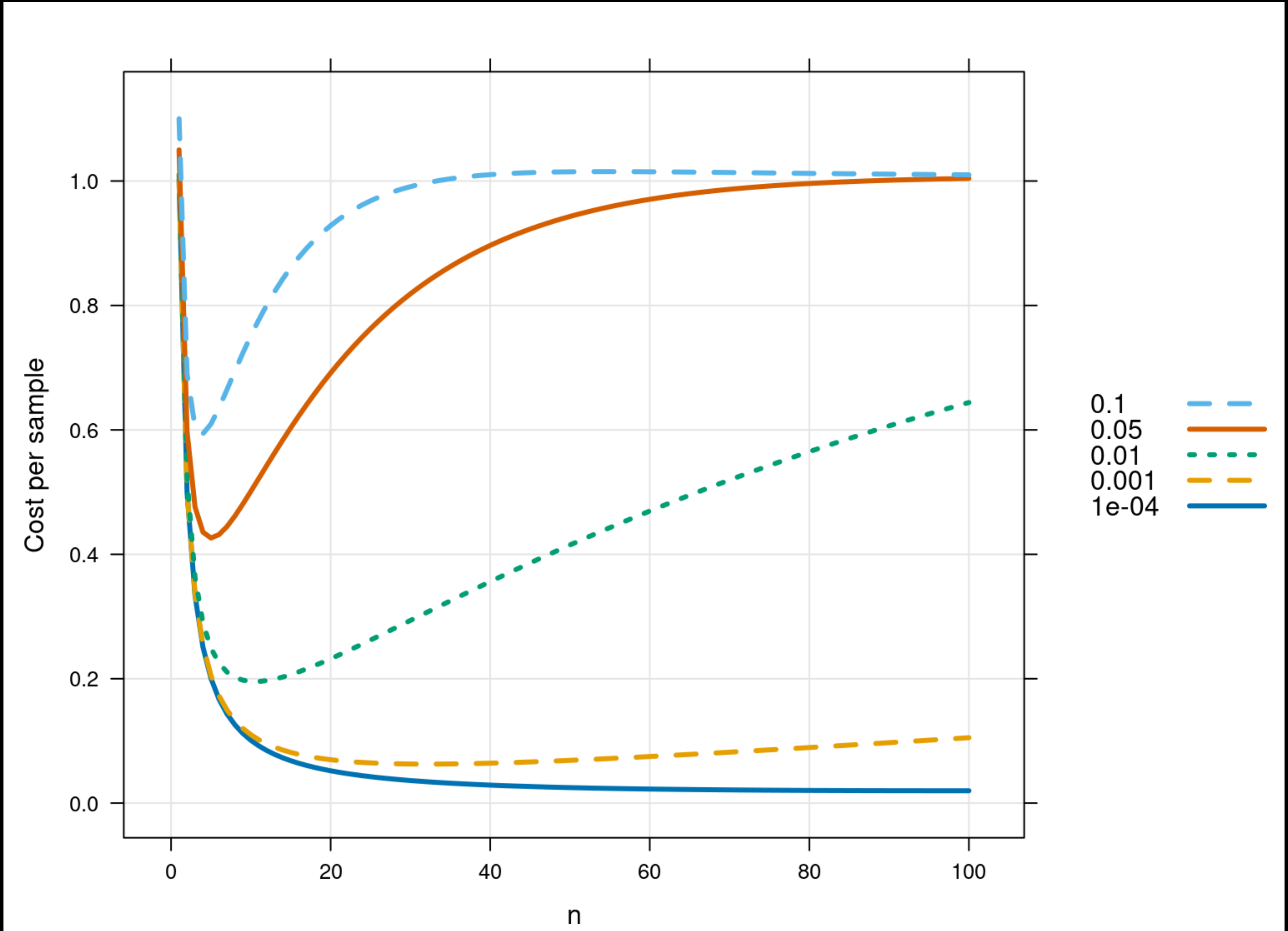
Then test.

If negative - done - all okay

If positive - test n samples.

Let p = probability of a positive.

$$E(\text{Tests}) = 1 \times \underbrace{(1-p)^n}_{\text{pr all negative}} + (n+1)\underbrace{\left(1-(1-p)^n\right)}_{\text{prob. at least 1 positive}}$$



Cost per sample
n
0.0
0.2
0.4
0.6
0.8
1.0
0
20
40
60
80
100
0.1
0.05
0.01
0.001
1e-04

```r
df <- expand.grid(n = 1:100, p = c(.0001,.001, .01, .05, .1))
head(df)
dim(df)
df <- within(df,
             {
               Etests <- 1* (1-p)^n + (n + 1) * (1 - (1 - p)^n)
               Cost_per_sample <- Etests/n
             }
             )

library(latticeExtra)
trellis.par.set(superpose.line = list(lwd=3, lty = 1:3))
xyplot(Cost_per_sample ~ n, df,
       groups = p, type = 'l',
       ylab = "Cost per sample",
       auto.key = list(reverse.rows = T)) +
  layer_(panel.grid(h=-1, v = -1))
```

Example D: Illustrates how

$$E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i)$$

does not require independence

DNA Sequences: formed from 4 letters A C T G

If random & each letter with = prob.

$$\underbrace{\text{A T C A A T C G A G T} \cdots \text{T A A}}$$

- Suppose length $= N$
- and each letter has $p = 1/4$

How many ATGC do you expect?

Let I_n = event ATGC starts at position n

$$P(I_n) = \frac{1}{4} \times \frac{1}{4} \times \frac{1}{4} \times \frac{1}{4} = \frac{1}{256}$$

$$= E(I_n)$$

if $I_n = \begin{cases} 1 & \text{if ATGC starts at position } n \\ 0 & \text{otherwise} \end{cases}$

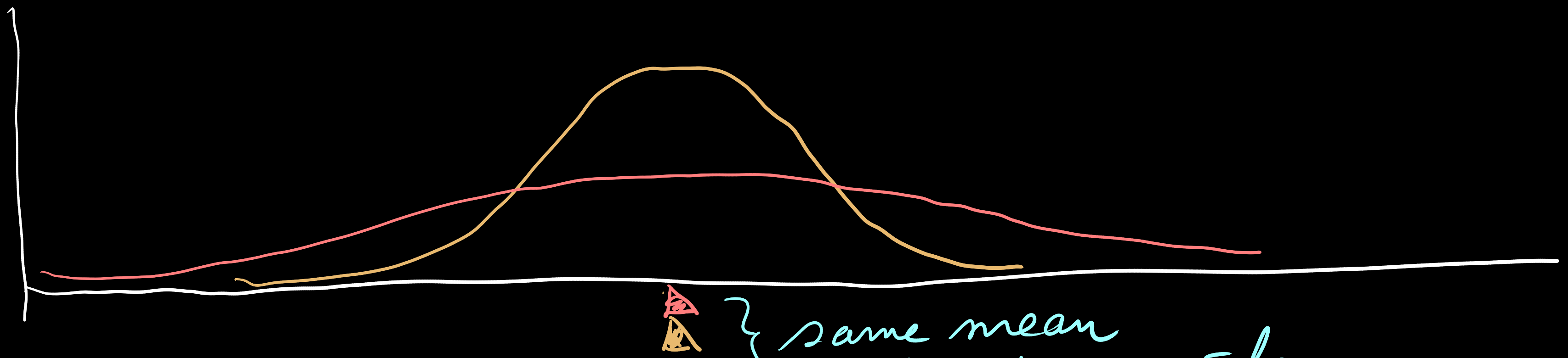
$$E(\#\text{of sequences}) = E\left(\sum_{n=1}^{N-3} I_n\right) = \sum_{n=1}^{N-3} E(I_n)$$

$$= (N-3) \times \frac{1}{256}$$

Variance and Standard Deviation

Mean = "location parameter" ← one of many e.g. median, min, max

Next we need a "spread" parameter



} same mean different spreads

Variance: Average squared distance from the mean

$$\mathrm{Var}(X) = E[(X-\mu_X)^2] \text{ if } E \text{ exists.}$$

in squared units

$$SD(X) = \sqrt{\mathrm{Var}(X)}$$

in original units

Facts about variance.

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

also a useful way to calculate Var(X)

Proof

$$
\begin{aligned}
&E[(X-\mu_X)^2] \\
&= E(X^2 - 2\mu_X X + \mu_X^2) \\
&= E(X^2) - 2\mu_X E(X) + \mu_X^2 \\
&= E(X^2) - 2E(X)E(X) + (E(X))^2 \\
&= E(X^2) - (E(X))^2
\end{aligned}
$$

Corollary: $E(X^2) = E(X)^2$ iff $Var(X) = 0$

Fact: $Var(X) = 0$ iff X is a constant

i.e. $P(X = c) = 1$

Fact: If $Var(X) < 0$ (i.e. exists)

and $Y = a + bX$

then $Var(Y) = b^2 Var(X)$

Chebyshev's Inequality: Prop & spread

Let X have mean μ and variance σ^2

Then for any $t > 0$:

$$P(|X-\mu| > t) \leq \sigma^2/t^2$$

Proof: Use Markov's inequality on $Y = (X-\mu)^2$

Different form: If $\sigma > 0$, let $t = k\sigma$ for $k > 0$. Then

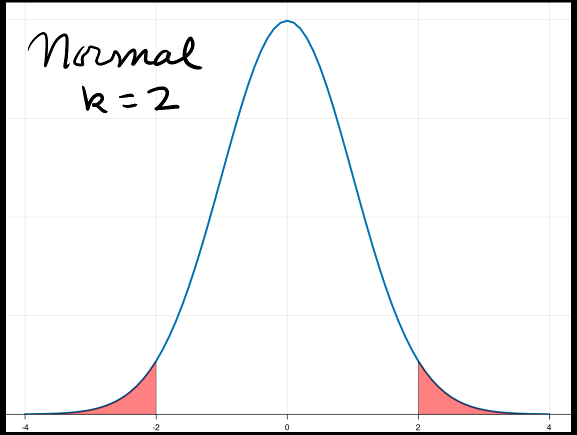
$$P(|X-\mu| > k\sigma) \leq \frac{\sigma^2}{k^2\sigma^2} = 1/k^2$$

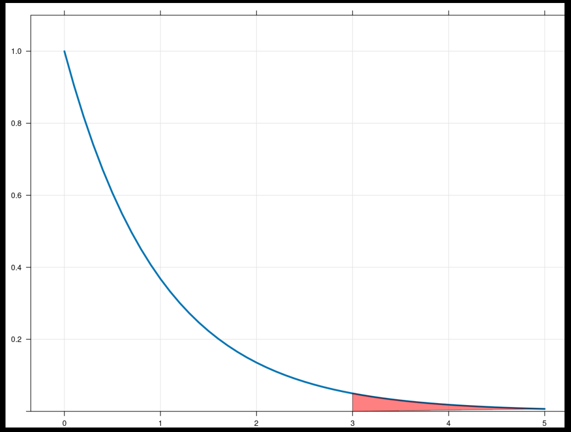
For $k = 1$ this says $P(|X-\mu| \geq \sigma) \leq 1/1^2 = 1$!?

Only useful for $k > 1$; e.g. $P(|X-\mu| \geq 2\sigma) \leq \frac{1}{2^2} = \frac{1}{4}$

Let's compare with some actual values.

	$k = 2$	$k = 3$	$k = 4$
Chebyshev	$\leq 1/4 = 0.25$	$\leq 1/9 = 0.1111\ldots$	$\leq \frac{1}{16} = 0.0625$
Actuals			
Normal	0.0455	0.0027	0.0000633
Exponential	0.0498	0.0183	0.00674
Poisson	0.0803	0.0190	0.00366





Fact (Corollary A, p. 134)

$$\mathrm{Var}(X) = 0 \quad \text{iff} \quad P(X=\mu) = 1$$
$$\text{iff} \quad P(X \text{ is constant}) = 1$$

Proof: Example of Chebyshev's Theorem.

$$P(|X-\mu| > \varepsilon) \leq \sigma^2/\varepsilon^2 = 0$$

for every $\varepsilon > 0$.

So $P(|X-\mu| = 0) = 1$

i.e. $P(X=\mu) = 1$

Investment Portfolios (p. 134)

Two investments: R_1: $E(R_1) = \mu_1 = 0.10$

$\sigma_1 = 0.075$

R_2: $E(R_2) = \mu_2 = 0.03$

$\sigma_2 = 0$

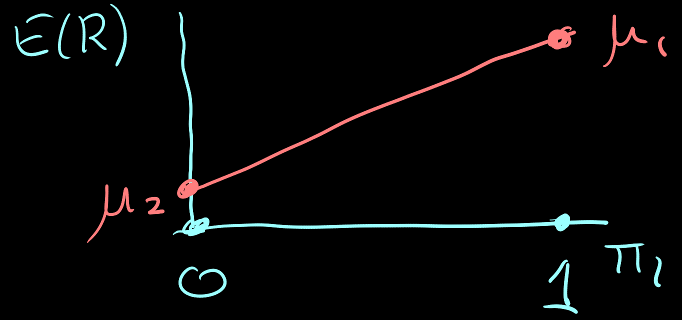
Same cost so investor can choose to invest $\underline{\pi_1 \text{ in } R_1}$ and $\underline{\pi_2 = (1-\pi_1) \text{ in } R_2}$

Return is a R.V.:

$$R = \pi_1 R_1 + (1-\pi_1) R_2$$

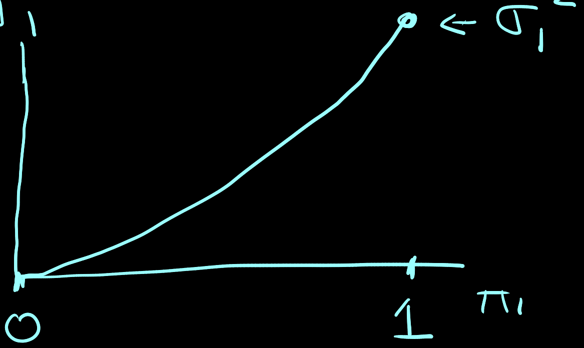
$$E(R) = \pi_1 \mu_1 + (1-\pi_1)\mu_2$$

Expected return



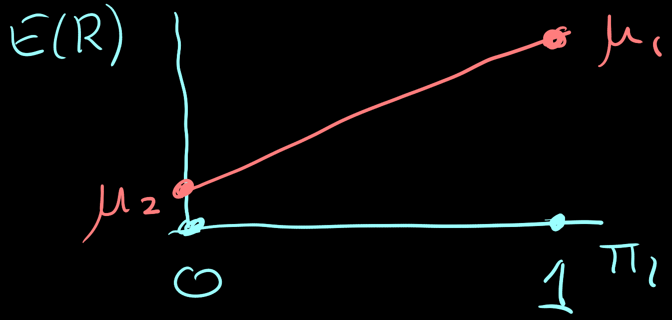
$$Var(R) = \pi_1^2 \sigma_1^2$$

Risk



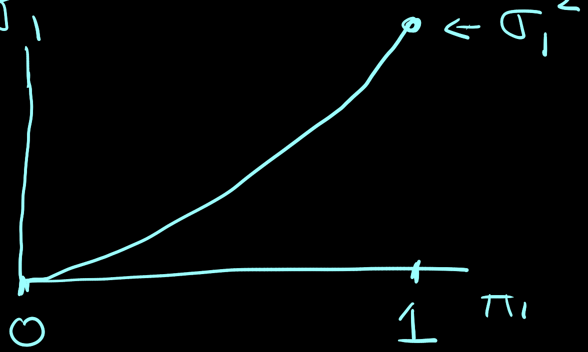
$$E(R) = \pi_1 \mu_1 + (1-\pi_1)\mu_2$$

Expected return



$$\mathrm{Var}(R) = \pi_1^2 \sigma_1^2$$

Risk



need to balance expected return against risk.

Notes: • When $\sigma_2 > 0$ we will also need to consider covariance

• Variance is <u>not</u> the whole story as far as risk is concerned.

Read 4.2.1 Measurement error

Suppose we want to measure some "fixed" but "unknown" quantity.

Call it "x_0"

Measurement $X = x_0 + \beta + \varepsilon$

fixed unknown

systematic error - same each time

random error varies with each measurement.

β = "bias"

ε = random component: $Var(\varepsilon) = \sigma^2$

$E(\varepsilon) = 0$

Then $\mathrm{Var}(X) = \mathrm{Var}(\underbrace{x_0 + \beta}_{\text{constant}} + \varepsilon)$

$$= \mathrm{Var}(\varepsilon) = \sigma^2$$

But $E(\text{Error}^2) = MSE$

$$= E\left[(X - x_0)^2\right]$$

$$= E\left[(\beta + \varepsilon)^2\right]$$

$$= E\left[\beta^2 + 2\beta\varepsilon + \varepsilon^2\right]$$

$$= \beta^2 + 2\beta \underbrace{E(\varepsilon)}_{=0} + E(\varepsilon^2)$$

$$= \mathrm{Var}(X) + (E(X))^2 = \mathrm{Var}(X) = \sigma^2$$

$$\text{"MSE"} = \beta^2 + \sigma^2 = \text{Bias}^2 + \sigma^2$$

4.3 Covariance & Correlation

$$\operatorname{Cov}(X, Y) = E\left[(X-\mu_X)(Y-\mu_Y)\right]$$

"with"

if E exists.

"times" is correct here

Fact: $$\operatorname{Cov}(X,Y) = E\left[XY - \mu_X Y - \mu_Y X + \mu_X \mu_X\right]$$

$$= E(XY) - \mu_X\mu_Y - \mu_Y\mu_X + \mu_X\mu_Y$$

$$= E(XY) - E(X)E(Y)$$

never say Covariance of X ~~times~~ Y

BUT always X "with" Y

Note: $\mathrm{Cov}(X, X) \stackrel{?}{=}$

Facts

$$\mathrm{Cov}(a+X,\, b+Y) = \mathrm{Cov}(X, Y)$$

$$\mathrm{Cov}(aX,\, bY) = ab\,\mathrm{Cov}(X, Y)$$

$$\mathrm{Cov}(X,\, Y+Z) = \mathrm{Cov}(X, Y) + \mathrm{Cov}(X, Z)$$

$$\mathrm{Cov}(aW+bX, cY+dZ)$$
$$= ac\,\mathrm{Cov}(W, Y) + ad\,\mathrm{Cov}(W, Z)$$
$$+ bc\,\mathrm{Cov}(X, Y) + bd\,\mathrm{Cov}(X, Z)$$

Summary:

Linearity of Expectation:

If $Y = a + b_1 X_1 + \cdots + b_n X_n$

then

$$E(Y) = a + b_1 E(X_1) + \cdots + b_n E(X_n)$$

In other words

If $Y = a + \sum_{i=1}^{n} X_i$

then $E(Y) = a + \sum_{i=1}^{n} E(X_i)$

Bilinearity of Covariance

$$\mathrm{Cov}\left(X, a + \sum_{i=1}^{n} b_i Y_i\right)$$

$$= \sum_{i=1}^{n} b_i \,\mathrm{Cov}(X, Y_i)$$

$$\mathrm{Cov}\left(a + \sum_{i=1}^{n} b_i X_i , Y\right)$$

$$= \sum_{i=1}^{n} b_i \,\mathrm{Cov}(X_i, Y)$$

$$\mathrm{Cov}\left(a+\sum_{i=1}^{n} b_i X_i \,,\, c+\sum_{j=1}^{m} d_j Y_j\right)$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{m} \underbrace{b_i d_j \,\mathrm{Cov}(X_i, Y_j)}_{\text{this should look like matrix multiplication}}$$