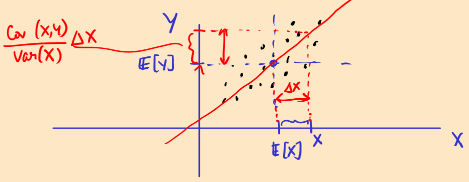
X = body fat %age

Y = prob of heart attack

| | | |
|---|---|---|
| | 0.05 | 15-20% |
| $E[Y \mid X = 23\%]$ | 0.06 | 20-25% |
| 23% | 0.08 | 25-30% |

Note 20

$$E[Y \mid X = x] = E[Y] + \underbrace{\frac{Cov(X,Y)}{Var(X)}}(x - E[X]) = L(Y \mid X)$$




**Cov is bilinear**

$\alpha, \beta \in \mathbb{R}$

1. $Cov(\alpha X, \beta Y)$
$= \alpha Cov(X, \beta Y)$
$= \beta Cov(\alpha X, Y)$
$= \alpha\beta Cov(X, Y)$

2. $Cov(X_1 + X_2, Y)$
$= Cov(X_1, Y) + Cov(X_2, Y)$

$Cov(X_1 + X_2, Y_1 + Y_2)$
$= Cov(X_1, Y_1)$
$+ Cov(X_1, Y_2)$
$+ Cov(X_2, Y_1)$
$+ Cov(X_2, Y_2)$

$Var(X) = Cov(X, X)$

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

$= E[(X - E[X])(X - E[X])]$

$= Cov(X, X)$

$Var(\alpha X) = Cov(\alpha X, \alpha X)$

$Cov(X, Y)$
$= Cov(Y, X)$

**ALWAYS TRUE**

$Var(X_1 + X_2)$

$= Cov(X_1 + X_2, X_1 + X_2)$

$= Cov(X_1, X_1)$
$+ Cov(X_2, X_2)$
$+ Cov(X_2, X_1)$
$+ Cov(X_1, X_2)$

$= Var(X_1) + Var(X_2) + 2Cov(X_1, X_2)$

If $Cov(X_1, X_2) = 0$

Then $Var(X_1 + X_2) = Var(X_1) + Var(X_2)$

1.(c) $Cov(X, Y)$

$= Cov(X_1 + X_2 + X_3, X_4 + X_5 + X_6)$

$= 9 \cdot Cov(X_1, X_4)$ bilinearity of Cov

$X_1, X_2, \ldots, X_6$ have same dist

$= 9 \cdot \frac{1}{144}$

$= \frac{9}{144}$

1.(d) $L(Y \mid X) = \frac{7}{4} + \underbrace{\frac{9/144}{111/144}\left(X - \frac{7}{4}\right)}$

out in

function of $x$.

input $x$

output : predicted value of $Y$.

# 1 LLSE

We have two bags of balls. The fractions of red balls and blue balls in bag $A$ are $2/3$ and $1/3$ respectively. The fractions of red balls and blue balls in bag $B$ are $1/2$ and $1/2$ respectively. Someone gives you one of the bags (unmarked) uniformly at random. You then draw 6 balls from that same bag with replacement. Let $X_i$ be the indicator random variable that ball $i$ is red. Now, let us define $X = \sum_{1 \le i \le 3} X_i$ and $Y = \sum_{4 \le i \le 6} X_i$.

(A) (B)

$\underbrace{X_1 \quad X_2 \quad X_3}_{X = \#\text{red}}$ $\quad$ $\underbrace{X_4 \quad X_5 \quad X_6}_{Y = \#\text{red}}$

(a) Compute $\mathbb{E}[X]$ and $\mathbb{E}[Y]$.

(b) Compute $\text{Var}(X)$.

(c) Compute $\text{cov}(X,Y)$. (*Hint*: Recall that covariance is bilinear.)

(d) Compute $L(Y \mid X)$, the best linear estimator of $Y$ given $X$. (*Hint*: Recall that

$$L(Y \mid X) = \mathbb{E}[Y] + \frac{\text{cov}(X,Y)}{\text{Var}(X)}(X - \mathbb{E}[X]).$$

)

$X_1$ has same dist as $X_2, X_3, \dots, X_6$

(a) $\mathbb{E}[X] = \mathbb{E}[X_1 + X_2 + X_3] = 3\,\mathbb{E}[X_1] = 3\Pr[X_1 = 1] = 3 \cdot \left(\frac{1}{2} \cdot \frac{2}{3} + \frac{1}{2} \cdot \frac{1}{2}\right) = \frac{7}{4}$

$\mathbb{E}[Y]$ //

$\frac{1}{3} + \frac{1}{4} = \frac{7}{12}$

| $X_2 \backslash X_1$ | 0 | 1 |
|---|---|---|
| 0 | 0 | 0 |
| 1 | 0 | 1 |

(b) $\text{Var}(X) = \text{Cov}(X, X)$

$= \text{Cov}(X_1 + X_2 + X_3, X_1 + X_2 + X_3)$

$= \sum_{1 \le i,j \le 3} \text{Cov}(X_i, X_j)$ bilinearity of Cov

$= \text{Cov}(X_1, X_1) + \text{Cov}(X_2, X_2) + \text{Cov}(X_3, X_3)$
$+ \text{Cov}(X_1, X_2) + \text{Cov}(X_1, X_3) + \text{Cov}(X_2, X_1)$
$+ \text{Cov}(X_2, X_3) + \text{Cov}(X_3, X_1) + \text{Cov}(X_3, X_2)$

$= 3\text{Var}(X_1) + 6\,\text{Cov}(X_2, X_1)$

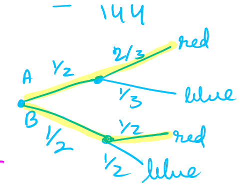
$= 3 \cdot \frac{35}{144} + 6 \cdot \frac{1}{144}$

$= \frac{111}{144}$

$\text{Cov}(X_1, X_2)$
$= \mathbb{E}[X_1 X_2] - \mathbb{E}[X_1]\mathbb{E}[X_2]$
$= \frac{50}{144} - \frac{49}{144} = \frac{1}{144}$

$\text{Var}(X_1)$
$= \mathbb{E}[X_1^2] - (\mathbb{E}[X_1])^2$
$= \mathbb{E}[X_1] - (\mathbb{E}[X_1])^2$
$= \frac{7}{12} - \frac{49}{144}$
$= 35/144$

$\mathbb{E}[X_1 X_2] = \Pr[X_1 X_2 = 1]$
$= \Pr[X_1 = 1]\Pr[X_2 = 1 \mid X_1 = 1]$
$= \frac{50}{144}$




| $X_i^2$ | $X_i$ | Pr |
|---|---|---|
| 0 | 0 | $1-p$ |
| 1 | 1 | $p$ |

$X_i$ is indicator
$X_i^2$ has same dist as $X_i$

# 2 Balls in Bins Estimation

We throw $n > 0$ balls into $m \ge 2$ bins. Let $X$ and $Y$ represent the number of balls that land in bin 1 and 2 respectively.

(a) Calculate $\mathbb{E}[Y \mid X]$. [*Hint*: Your intuition may be more useful than formal calculations.]

(b) What is $L[Y \mid X]$ (where $L[Y \mid X]$ is the best linear estimator of $Y$ given $X$)? [*Hint*: Your justification should be no more than two or three sentences, no calculations necessary! Think carefully about the meaning of the conditional expectation.]

(c) Unfortunately, your friend is not convinced by your answer to the previous part. Compute $\mathbb{E}[X]$ and $\mathbb{E}[Y]$.

(d) Compute $\text{Var}(X)$.

(e) Compute $\text{cov}(X,Y)$.

(f) Compute $L[Y \mid X]$ using the formula. Ensure that your answer is the same as your answer to part (b).

# 3 Continuous LLSE

Suppose that $X$ and $Y$ are uniformly distributed on the shaded region in the figure below.

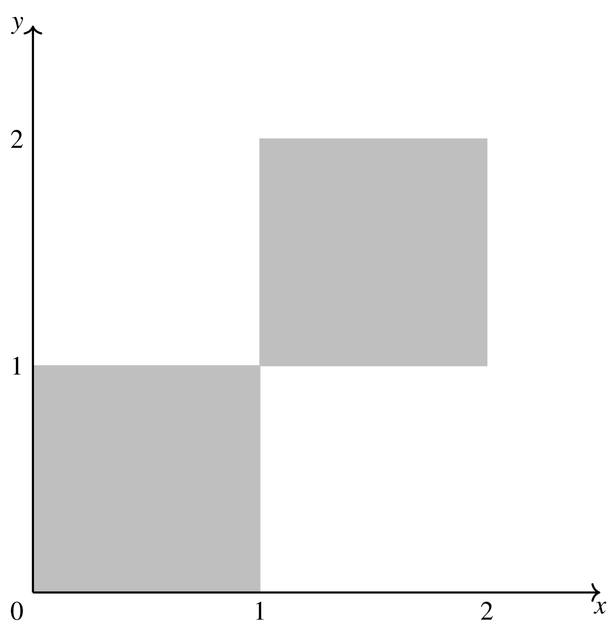



Figure 1: The joint density of $(X,Y)$ is uniform over the shaded region.

That is, $X$ and $Y$ have the joint distribution:

$$f_{X,Y}(x,y) = \begin{cases} 1/2, & 0 \le x \le 1, 0 \le y \le 1 \\ 1/2, & 1 \le x \le 2, 1 \le y \le 2 \end{cases}$$

(a) Do you expect $X$ and $Y$ to be positively correlated, negatively correlated, or neither?

(b) Compute the marginal distribution of $X$.

(c) Compute $L[Y \mid X]$, the best linear estimator of $Y$ given $X$.

(d) What is $\mathbb{E}[Y \mid X]$?