L3 Supp. Material

Fair Dice example

| | | |
|---|---|---|
| 1 | 2 | 3 |
| 4 | 5 | 6 |

Sample space. $\Omega$

$\{1, 2, 3, 4, 5, 6\}$.

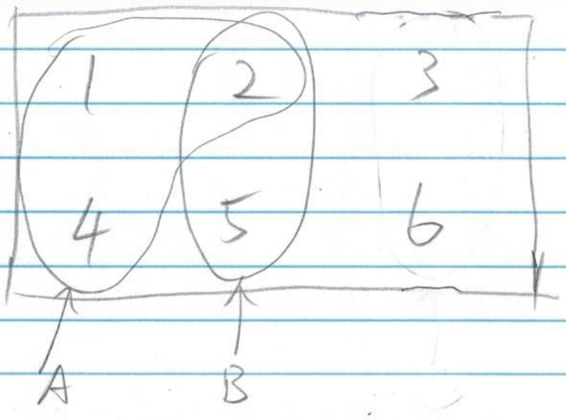
Event space.

$$F = \left\{ \begin{array}{l} \{1\}, \{2\}, \cdots \\ \{1,2\}, \{1,3\}, \cdots \\ \vdots \\ \{1,2,\cdots,5,6\}, \{\} \end{array} \right\}$$

Probability measure:

$$P: \left\{ \begin{array}{l} \{1\} \xrightarrow{P} \frac{1}{6} \\ \{2\} \longrightarrow \frac{1}{6} \\ \vdots \\ \{6\} \longrightarrow \frac{1}{6} \end{array} \right.$$

## Fair Dice Cont'd




$\{1,2,4\}$ $\{2,5\}$

$A \cup B = \{1,2,4,5\}$

$A \cap B = \{2\}$

$\Omega \setminus B = \{1,3,4,6\}$.

$P(A) = \frac{3}{6} = \frac{1}{2}$

$P(B) = \frac{2}{6} = \frac{1}{3}$

$P(A \cup B) = \frac{4}{6} = \frac{2}{3}$

$< P(A) + P(B)$

$P(\Omega \setminus B) = \frac{4}{6} = \frac{2}{3}$

$= 1 - P(B)$

$P(A \cap B) = \frac{1}{6}$

---

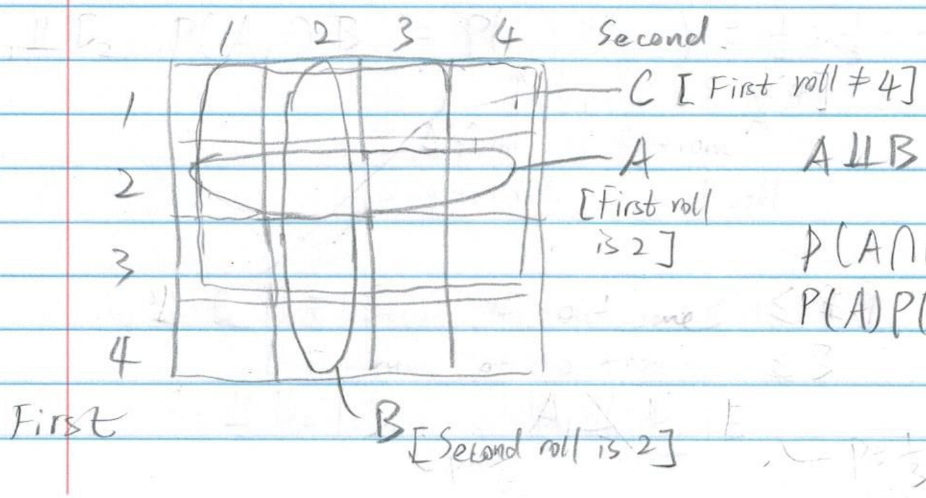
$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{1/6}{1/3} = \frac{1}{2}$.


B as new Sample space

---

Independence

4-face Dice, twice:




$A \perp\!\!\!\perp B$, as

$P(A \cap B) = \frac{1}{16}$

$P(A)P(B) = \frac{1}{4} \times \frac{1}{4} = \frac{1}{16}$

$A \perp\!\!\!\perp B \mid C$ as

$$P(A \cap B \mid C) = \frac{P(A \cap B \cap C)}{P(C)} = \frac{1/16}{12/16} = \frac{1}{12}$$

$$P(A \mid C)\, P(B \mid C) = \frac{4}{12} \cdot \frac{3}{12} = \frac{1}{12}.$$

Let $D = [\text{Sum of two rolls} \leq 4]$

$A \overset{?}{\perp\!\!\!\perp} B \mid D$

---

Bayes' Rule.

Screen test. $B_1$ – Infected, $B_2$ – Healthy
$A_1$ – Positive ; $A_2$ – Negative

1% Infected — 90% Positive
— 10% False negative

99% Healthy — 5% False positive
— 95% Negative

$$P(B_1 \mid A_1) = \frac{P(A_1 \mid B_1)\, P(B_1)}{P(A_1)}, \quad P(A_1) = P(A_1 \mid B_1) P(B_1) + P(A_1 \mid B_2) P(B_2)$$

$$= 0.9 \cdot 0.01 + 0.05 \cdot 0.99$$

$$\approx 0.058$$

$$P(B_1 \mid A_1) = \frac{0.9 \cdot 0.01}{0.058}$$

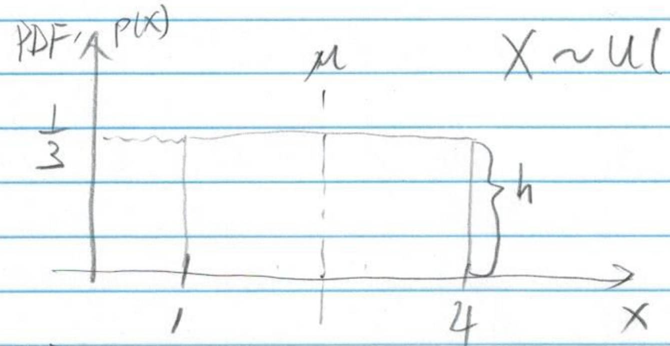
$$\approx 15.4\%$$

Positive test means 84.6% infection.

Example.

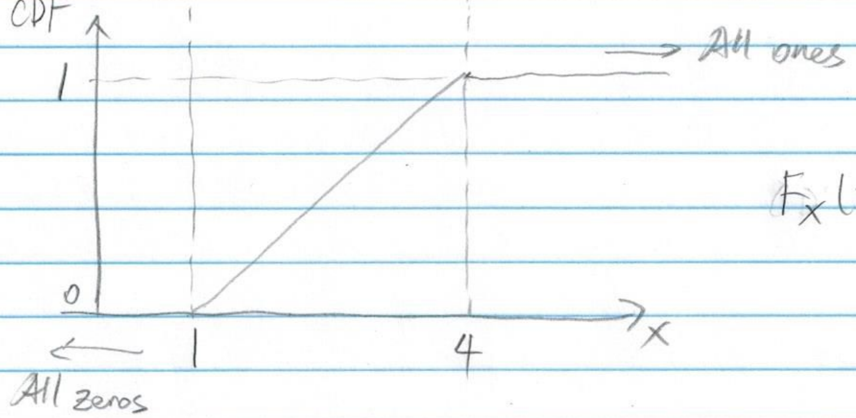
Uniform dist.

$X \sim U(1,4)$




$$1 = \int_{-\infty}^{\infty} p(x)\,dx = \int_1^4 h\,dx = 3h$$

$$\Rightarrow h = \frac{1}{3}$$




$$F_X(x) = \begin{cases} 0, & x \le 1 \\ \frac{1}{3}(x-1), & x \in (1,4] \\ 1, & x > 4 \end{cases}$$

$$P(2 < x \le 3) = \int_2^3 p(x)\,dx = F_X(3) - F_X(2) = \frac{1}{3}$$

Expectation:

$$E[X] = \int_{-\infty}^{\infty} x\,p(x)\,dx = \int_1^4 x \cdot \frac{1}{3}\,dx = \frac{5}{2}$$

Variance:

$$E[(X - E[X])^2] = \int_1^4 \frac{1}{3}\left(x - \frac{5}{2}\right)^2 dx = \frac{3}{4}$$

Alt.

$$E[X^2] = \int_1^4 \frac{1}{3} x^2\,dx = 7$$

$$Var = E[X^2] - E[X]^2 = 7 - \left(\frac{5}{2}\right)^2 = \frac{3}{4}$$

## Covariance

$$Cov[X,Y] = E[(X-EX)(Y-EY)]$$
$$= E[XY - (EX)Y - X(EY) + (EX)(EY)]$$
$$= E[XY] - (EX)(EY) - \cancel{(EX)(EY)} + \cancel{(EX)(EY)}$$

$$Var[X+Y] = E[(X+Y-EX-EY)^2]$$
$$= E[(X-EX)^2 + (Y-EY)^2 + 2(X-EX)(Y-EY)]$$
$$= Var[X] + Var[Y] + 2Cov[X,Y]$$

---

## Likelihood example

since i.i.d.

$$P(D|w) = P(\{x_1, x_2, \ldots, x_N\}|w)$$ Same w, since i.d.

$$= P(x_1|w)P(x_2|w) \cdots P(x_N|w)$$ Decompose, since independent

$$= \prod P(x_i|w) \quad , \quad w = \{\mu, \sigma^2\}$$

$$p(x_i|w) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x_i-\mu)^2}{2\sigma^2}\right]$$

$$\text{Max } P(D|w) \Leftrightarrow \text{Min } -\log P(D|w)$$

$$L = -\log P(D|w) = -\sum_i \log p(x_i|w)$$

$$= \frac{1}{2\sigma^2}\left[\sum_i (x_i-\mu)^2\right] + N\log\sqrt{2\pi} + \frac{N}{2}\log\sigma^2$$

Reason for logarithm

* Easier derivative
* Avoid underflow.
  e.g. $10^{-18} \to 18$

At minimizer.

$$\frac{\partial L}{\partial \mu} = 0 = -\frac{1}{\sigma^2}\sum_i (x_i - \mu)$$

$$\Rightarrow \sum_i (x_i - \mu) = 0 \Rightarrow \mu_{ML} = \frac{1}{N}\sum_i x_i \rightarrow \text{Statistical mean.}$$

$$\frac{\partial L}{\partial \sigma^2} = 0 = -\frac{1}{2(\sigma^2)^2}\sum_i (x_i - \mu_{ML})^2 + \frac{N}{2}\frac{1}{\sigma^2}$$

$$\Rightarrow \sigma^2_{ML} = \frac{1}{N}\sum_i (x_i - \mu_{ML})^2 \rightarrow \text{Statistical variance (Biased).}$$

$$E[\sigma^2_{ML}] = \frac{N-1}{N}\sigma^2 \neq \sigma^2$$

Reason for bias : $\mu$ is dependent on $\{X_i\}$. so degrees-of-freedom of $\sigma^2_{ML}$ is $N-1$ instead of $N$.

Unbiased estimate is $\sigma^2 = \frac{1}{N-1}\sum_i (x_i - \mu)^2$ ← Bessel correction

Proof: $\sigma^2_{ML} = \frac{1}{N}\sum_i [(x_i - \mu) - (\mu_{ML} - \mu)]^2 = \frac{1}{N}\left[\sum_i (x_i - \mu)^2 - N(\mu_{ML} - \mu)^2\right]$

$$E[\sigma^2_{ML}] = \frac{1}{N}\left[\sum_i E(x_i - \mu)^2 - N \underbrace{E(\mu_{ML} - \mu)^2}_{\text{II}}\right]$$

$$= \frac{1}{N}\left[N \times \sigma^2 - N \cdot \frac{1}{N}\sigma^2\right]$$ Left as exercise

$$= \frac{N-1}{N}\sigma^2$$

# A more Bayesian ML example.

$\Delta$ Want to determine a parameter $\mu$ in a model

1. Before measurement we have a belief, i.e. prior. $\mu \sim N(\mu_0, \sigma_0^2)$ known

2. Now measurement of the model is $x \sim N(x_0, \sigma^2)$

   The probability of getting $x_0$ given $\mu$ is the likelihood,

$$p(x_0|\mu) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(\mu-x_0)^2}{2\sigma^2}\right]$$

i.e. when truth is $\mu$, the prob. of getting $x_0$ in measurement.

3. The measurement updates our belief about $\mu$, i.e. posterior

$$p(\mu|x_0) = \frac{p(x_0|\mu)\,p(\mu)}{p(x_0)} \propto p(x_0|\mu)\,p(\mu)$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(\mu-x_0)^2}{2\sigma^2}\right] \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left[-\frac{(\mu-\mu_0)^2}{2\sigma_0^2}\right]$$

Cheating; for Gaussian prior & likelihood, posterior is Gaussian.

Assume $p(\mu|x_0) \propto \exp\left[-\frac{(\mu-\mu_1)^2}{2\sigma_1^2}\right]$, only need to determine $\mu_1, \sigma_1^2$.

$$\propto \exp\left(-\frac{\mu^2}{2}\boxed{\frac{1}{\sigma_1^2}} + \mu \boxed{\frac{\mu_1}{\sigma_1^2}}\right)$$

$\exp\left(-\frac{\mu_1^2}{2\sigma_1^2}\right)$ is const

$$\propto \exp\left[-\frac{\mu^2}{2}\boxed{\left(\frac{1}{\sigma_0^2}+\frac{1}{\sigma^2}\right)} + \mu\boxed{\left(\frac{\mu_0}{\sigma_0^2}+\frac{x_0}{\sigma^2}\right)}\right]$$

So $\sigma_1^2 = (\sigma_0^{-2} + \sigma^{-2})^{-1}$, $\mu_1 = \sigma_1^2(\mu_0\sigma_0^{-2} + x_0\sigma^{-2})$

Old guess $(\mu_0, \sigma_0^2)$ updated to $(\mu_1, \sigma_1^2)$.

New confidence level.

New guess

Comment: $p(x_0) = \int p(x_0|\mu)\,p(\mu)\,d\mu$ is called evidence.

$p(x_0)$ depends on the form of $p(x_0|\mu)$, i.e. the model.

Good model $\leftrightarrow$ High $p(x_0)$, hence the name evidence.

or Good measurement in this particular example