**Topics Covered**

# I Random Variables

- $X_1, ..., X_n$ are independent if $P(X_1 \in A_1 \overset{X}{\underset{n}{\dots}} \in A_n) = \prod_{i=1}^{n} P(X_i \in A_i)$
- IID - independent identically distributed variables $X_1, ..., X_n$ IID $\rightarrow \langle X_1, ..., X_n \rangle$ is a *random sample of size n.*
- A function of an RV is also an RV.
- Sample mean $\bar{X} = \dfrac{X_1 + ... + X_n}{n}$

Assime $X_1, ..., X_n$ are IID, with

$$\begin{aligned} E(X_i) &= \mu \\ V(X_i) &= E(X - E(X))^2 = \sigma^2 \end{aligned}$$

**Then:**

$$\begin{aligned} E(\bar{X}) &= \mu \\ V(\bar{X}) &= \frac{\sigma}{n} \end{aligned}$$

Evaluating with the sample mean gives:

$$\begin{aligned} E(\bar{X}) &= \frac{nE(X_1)}{n} \\ &= E(X_1) \end{aligned}$$

$$\begin{aligned} E(\bar{X} - \mu)^2 &= E(\frac{\sum_{i=1} n(X_i - \mu)}{n})^2 \\ &= \frac{1}{n^2}(\sum_{i=1} n(X_i - \mu)^2 + \sum_{i \neq j} n(X_i - \mu)(X_j - \mu)) \\ &= \frac{1}{n^2} \cdot n\sigma^2 \\ &= \frac{\sigma^2}{n} \end{aligned}$$

So $\bar{X}$ is an unbiased estimator of the true value (a.k.a. the population mean). What is an unbiased estimator of the variance $\sum_i (X_i - \bar{X})^2$?

$$
\begin{aligned}
E(\sum_i (X_i - \bar{X})^2) &= E(\sum_i (X_i^2 - 2X_i\bar{X} + \bar{X}^2)) \\
&= E(\sum_i X_i^2 - 2(\sum_i X_i)\bar{X} + n\bar{X}^2) \\
&= E(nX^2 - 2n\bar{X}^2 + n\bar{X}^2) \\
&= nE(nX^2) - nE(\bar{X}^2)
\end{aligned}
$$

What is $E(X^2)$?

$$
\begin{aligned}
\sigma^2 &= E(X - \mu)^2 \\
&= E(X^2 - 2\mu X + \mu^2) \\
&= E(X^2) - 2\mu^2 + \mu^2) \\
&= E(X^2) - \mu^2)
\end{aligned}
$$

So

$$E(X^2) = \sigma^2 + \mu^2 \tag{1}$$

So we get

$$
\begin{aligned}
E(\sum_i (X_i - \bar{X})^2) &= n \cdot (\sigma^2 + \mu^2) - n \cdot (\frac{\sigma^2}{n} - \mu^2) \\
&= n\sigma^2 + n\mu^2 - \sigma^2 - n\mu^2 \\
&= n\sigma^2 + \sigma^2 \\
&= (n-1)\sigma^2
\end{aligned}
$$

So an unbiased variance estimator is

$$\frac{1}{n-1} \sum_i (X_i - \bar{X})^2 \tag{2}$$

# II Likelihood

Assume you have a coin with $P(H) = p$ which is unknown. You toss it 16 times and get

$$HHTHTHHHTTHHTHHT$$

For what value of $p$ is such an outcome *most likely*? We have 10 heads so intuitively $p$ should be $\frac{10}{16}$. To show this, start with

$$\begin{aligned}
P(p) &= p^{10}(1-p)^6 \\
\frac{\partial P}{\partial p} &= 10p^9(1-p)^6 + p^{10} \cdot 6(1-p)^5(-1) \\
&= p^9(1-p)^5(10(1-p) - 6p) \\
&= p^9(1-p)^5(10 - 10p - 6p)
\end{aligned}$$

So $P(p)$ has a maximum value

$$\begin{aligned}
16p &= 10 \\
p &= \frac{10}{16}
\end{aligned}$$

The likelihood function is **not** a probability function on the space of parameters. It is equal to:

1. In the discrete case: The probability to have such an outcome given the unknown parameters.
2. In the continuous case: The probability density function for such an outcome.

# Example 1

Assume you have $n$**unbiased** sensors with an unknown but equal standard deviation. You have $n$ readings $X_1, ..., X_n$ of the same quantity. How would you estimate the true value $\mu$ and $\sigma$?

Example 1 continued on next page... 

$$\begin{aligned}
\mathcal{L}_n(\mu,\sigma) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(X_i-\mu)^2}{2\sigma^2}} \\
\frac{\partial \mathcal{L}_n}{\partial \mu} &= \frac{\partial}{\partial \mu}\left(\frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} e^{-\frac{\sum_i (X_i-\mu)^2}{2\sigma^2}}\right) \\
&= \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} e^{-\frac{\sum_i (X_i-\mu)^2}{2\sigma^2}} \cdot \frac{-2}{2\sigma^2} \sum_i (X_i-\mu) \text{ because } \mu = \frac{\sum_i X_i}{n} \\
&= -n(2\pi)^{\frac{-n}{2}} \sigma^{-n-1} e^{-\frac{\sum_i (X_i-\mu)^2}{2\sigma^2}} + (2\pi)^{\frac{-n}{2}} \sigma^{-n} e^{-\frac{\sum_i (X_i-\mu)^2}{2\sigma^2}} (-2) \frac{-\sum_i (X_i-\mu)^2}{2\sigma^3}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial \mathcal{L}_n}{\partial \mu} &= 0 \leftrightarrow -n + (\sigma^{-1})^2 = 0 \\
&\leftrightarrow \sum_i (X_i -)^2 = 0 \\
\sigma^2 &= \frac{\sum_i (X_i -)^2}{n}
\end{aligned}$$

We found that $\mu = \bar{X}$, so

$$\sigma^2 = \frac{\sum_i (X_i -)^2}{n}$$

is a biased estimator.

# Example 2

Given $n$ sensors with known variances $\sigma_1^2, ..., \sigma_n^2$ and $n$ readings $X_1, ..., X_n$. What is the maximum likelihood estimation of $\mu$?

$$\begin{aligned}
\mathcal{L}_n(\vec{X},\mu,\sigma) &= \prod_{i=1}^{n} \frac{1}{(2\pi)^{\frac{n}{2}} \sigma_i} e^{-\frac{(X_i-\mu)^2}{2\sigma^2}} \\
&= (2\pi)^{-\frac{n}{2}} \prod_{i=1}^{n} \sigma_i^{-1} \cdot e^{-\sum_i \frac{(X_i-\mu)^2}{2\sigma^2}} \\
\frac{\partial \mathcal{L}_n}{\partial \mu} &= (2\pi)^{-\frac{n}{2}} \prod_{i=1}^{n} \sigma_i^{-1} \cdot e^{-\sum_i \frac{(X_i-\mu)^2}{2\sigma^2}} - 2\sum_i \frac{X_i-\mu}{2\sigma_i}
\end{aligned}$$

Example 2 continued on next page. . . 

So

$$\frac{\partial \mathcal{L}_n}{\partial \mu} = 0 \leftrightarrow \sum_i \frac{X_i}{\sigma_i} = \sum_i \frac{1}{\sigma_i}\mu$$
$$\leftrightarrow \mu = \frac{\sum_i \frac{1}{\sigma_i} X_i}{\sum_k \frac{1}{\sigma_k}}$$

# III Maximum Likelihood Estimates

Is the ML estrimate always a "good" one?

# Example 1

You have a box with a certain number of balls, numbered consecutively $1, 2, 3, 4, \ldots$. You pick one at random, see its number and have to estimate the total number of balls in the box.
**Maximum Likelihood Estimate**: If you picked a ball with a number $k$, such a ball is most likely if there are $k$ balls in the box, because then the probability of picking this ball is $\frac{1}{k}$. The expected value of your estimate is

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

So your ML is heavily biased! On the other hand, if your estimate is $2X - 1$ then

$$E(2X-1) = \sum_i \frac{2 \cdot i - 1}{n}$$
$$= \frac{2n(n+1) - n}{n}$$
$$= n$$

hence this estimate is unbiased.

Then to evaluate

$$T_i = C - \frac{1}{m_i} \sum_{i \to p} (E_{ip} - \sqrt{p})^2$$

We maximise

Example 1 continued on next page. . . 

$$\begin{aligned}
\tau &= \sum_i m_i T_i^2 \\
&= \sum_i m_i (C - \frac{1}{m_i} \sum_{i \to p} (E_{ip} - \sqrt{p})^2)^2 \\
\frac{\partial \tau}{\partial \sqrt{k}} &= 4 \sum_{i \to k} m_i (C - \frac{1}{m_i} \sum_{i \to p} (E_{ip} - \sqrt{p})^2)^2 \cdot \frac{E_{ik} - \sqrt{k}}{m_i} \\
&= 4 \sum_{i \to k} T_i (E_{ik} - \sqrt{k}) \\
&= 4 \sum_{i \to k} T_i E_{ik} - (\sum_{i \to k} T_i) \sqrt{k}
\end{aligned}$$

Equating the derivative to zero

$$\frac{\partial \tau}{\partial \sqrt{k}} = 0 \leftrightarrow \sqrt{k} = \frac{\sum_{i \to k} T_i E_{ik}}{\sum_{i \to k} T_i}$$

# Example 2

What is the likelihood of obtaining readings $E_{ip}$ if all sensors have variance $\sigma^2$?

$$\begin{aligned}
\mathcal{L}_n(E^i, \vec{r}, \sigma) &= \prod_{i \to k} \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(E_{ik} - \sqrt{k})^2}{2\sigma}} \\
ln(\mathcal{L}_n(E^i, \vec{r}, \sigma)) &= \sum_{i \to k} -\frac{1}{2} ln(2\pi\sigma) - \frac{(E_{ik} - \sqrt{k})^2}{2\sigma} \\
&= -\frac{1}{2} ln(2\pi\sigma) \cdot m_i - \frac{\sum_{i \to k} (E_{ik} - \sqrt{k})^2}{2\sigma} \\
\frac{2\sigma^2}{m_i} ln(\mathcal{L}_n(E^i, \vec{r}, \sigma)) &= -\sigma^2 ln(2\pi\sigma) - \sum_{i \to k} (E_{ik} - \sqrt{k})^2 \\
\frac{2\sigma^2}{m_i} ln(\mathcal{L}_n(E^i, \vec{r}, \sigma)) &= [-\sigma^2 ln(2\pi\sigma) - C] + [C - \sum_{i \to k} (E_{ik} - \sqrt{k})^2]
\end{aligned}$$

where $C$ is chosen such that $[-\sigma^2 ln(2\pi\sigma) - C] = 0$. Let $T_i = C - \sum_{i \to k} (E_{ik} - \sqrt{k})^2$, and so

Example 2 continued on next page...

$$
\begin{aligned}
(\frac{2\sigma^2}{m_i} ln(\mathcal{L}_n(E^i, \vec{r}, \sigma)))^2 &= T_i^2 \\
\frac{2\sigma^4}{m_i}(ln(\mathcal{L}_n(E^i, \vec{r}, \sigma)))^2 &= m_i T_i^2 \\
\sum_i m_i T_i^2 = 2\sigma^4 \sum_i \frac{ln\mathcal{L}_i(E^i, \vec{r}, \sigma)^2}{m_i}
\end{aligned}
$$

Let again

$$
g(\vec{r}) = \sum_i m_i T_i^2
$$

then

$$
\begin{aligned}
\frac{\partial g(\vec{r})}{\partial \sqrt{k}} &= 2 \sum_i m_i T_i \frac{\partial T_i}{\partial \sqrt{k}} \\
&= 2 \sum_i m_i T_i \frac{\partial (C - \sum_{p \to k}(E_{ip} - \sqrt{p})^2)}{\partial \sqrt{k}} \\
&= -4 \sum_i m_i T_i (-\frac{1}{m_i}) \sum_{fixme} (E_{ik} - \sqrt{k}) \\
&= 4 \sum_i T_i \sum_{fixme} (E_{ik} - \sqrt{k}) \\
&= 4 \sum_i T_i E_{ik} - 4 \frac{\sum_{i \to k}(T_i \sqrt{k})}{i \sum_{fixme} T_i} \\
\frac{1}{4 \sum_{i \to k} T_i} \frac{\partial g(\vec{r})}{\partial \sqrt{k}} &= \sum_i T_i E_{ik} - \sqrt{k} \\
r_k^{t+1} &= r_k^t + \frac{1}{4 \sum_{i \to k} T_i} \frac{\partial g(r)}{\partial r_k} \\
r^{t+1} &= \vec{r} + \frac{1}{4 \sum_{i \to k} T_i} \nabla g
\end{aligned}
$$