Review

RV is function $\Omega \to \mathbb{R}$

Distribution of RV X is described by

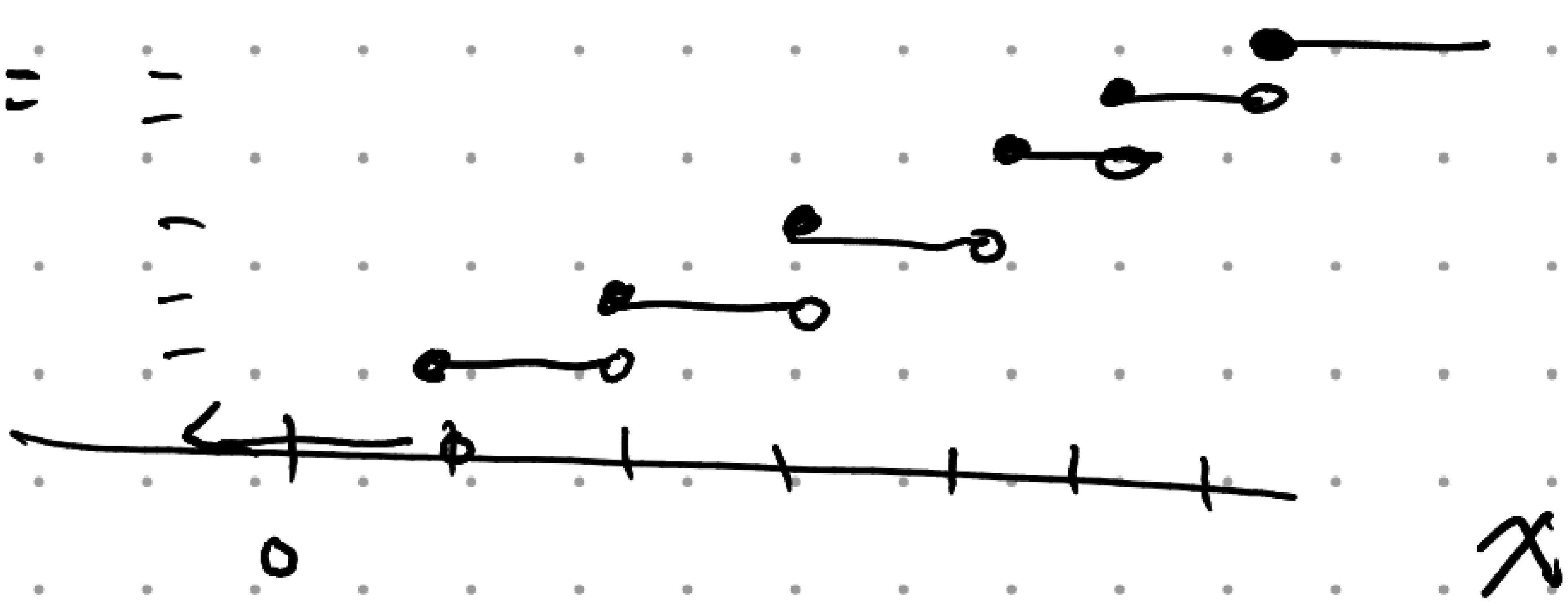
$F(x) = \mathbb{P}[X \le x]$

$\uparrow$ local var

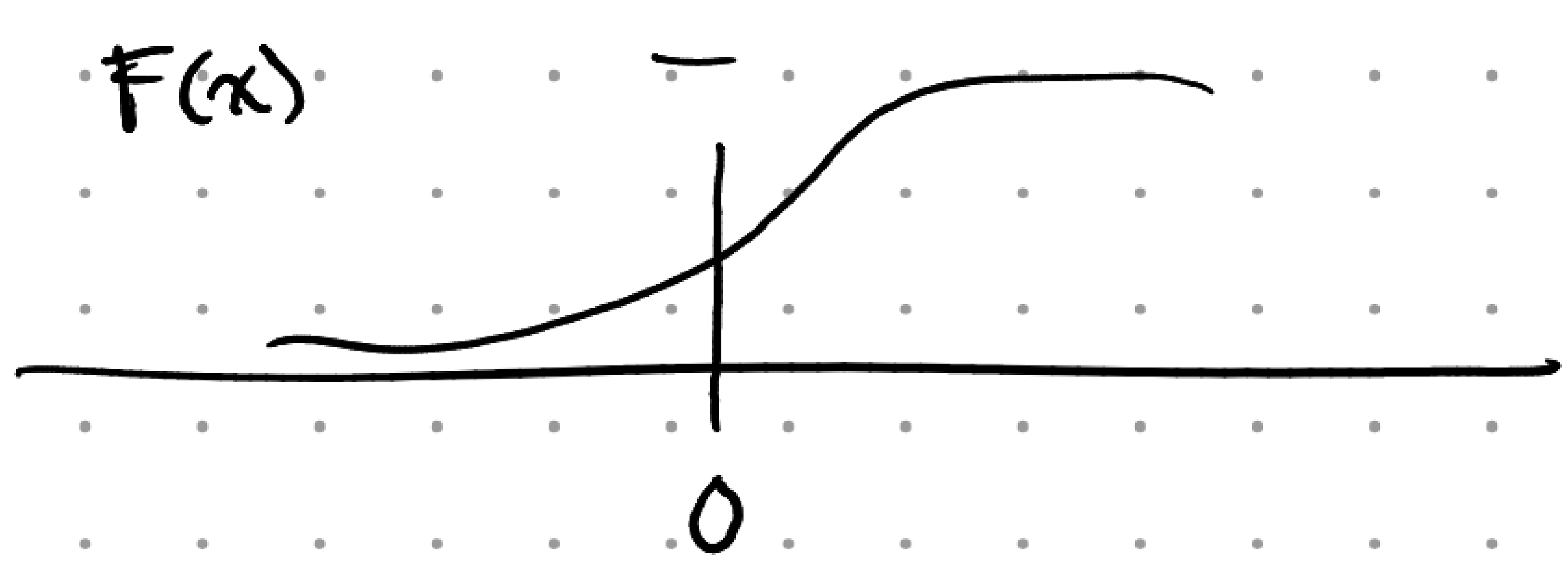
$$\left(\begin{array}{l} F(t) = \mathbb{P}[X \le t] \\ F_X(x) = \mathbb{P}[X \le x] \\ F_X(s) = \mathbb{P}[X \le s] \end{array} \right)$$

Ex. $X = \begin{cases} 1 & \text{w.p. } 1/6 \\ 2 & \vdots \\ 3 & \vdots \\ 4 & \\ 5 & \\ 6 & \end{cases}$

$F(x) =$



Ex. X standard normal



In general F a CDF if

- $F(b) \geq F(a)$ if $b \geq a$
- $\lim_{x \to \infty} F(x) = 1$ $\lim_{x \to -\infty} F(x) = 0$

$$\mathbb{P}[X \in (a, b]] = F(b) - F(a)$$

If X is discrete (values X can take are discrete)

pmf: $f(x) = \mathbb{P}[X = x] = \lim_{c \to x} F(x) - F(c)$

If X is continuous, $\mathbb{P}[X = x] = 0 \ \forall x$.

pdf: $f(x) = F'(x) = \lim_{dx \to 0} \dfrac{F(x+dx) - F(x)}{dx}$

$$= \lim_{dx \to 0} \frac{\mathbb{P}[X \in (x, x+dx]]}{dx}$$

$$\mathbb{E}[r(X)] = \int r(x)\, dF(x) = \begin{cases} \int r(x) \overset{\text{pdf}}{f(x)}\, dx & X \text{ continuous} \\ \sum_x r(x) \overset{\text{pmf}}{f(x)} & X \text{ discrete} \end{cases}$$

Ex. $F(x) = \begin{cases} 0 & x < 1 \\ 1/6 & 1 \le x < 2 \\ 2/6 & 2 \le x < 3 \\ 3/6 & 3 \le x < 4 \\ 4/6 & 4 \le x < 5 \\ 5/6 & 5 \le x < 6 \\ 1 & 6 \le x \end{cases}$

pmf $f(x) = \begin{cases} 1/6 & x = 1 \\ 1/6 & x = 2 \\ & \vdots \\ 1/6 & x = 6 \end{cases}$

$$\int x \, dF(x) = \sum_x x f(x) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + \cdots + 6 \cdot \frac{1}{6}$$

$$= \frac{7}{2}$$

$\int (x-\mu)^2 \, dF(x)$, $\mu = \int x \, dF(x)$

$$= \sum_x (x - 3.5)^2 f(x) = (1-3.5)^2 \cdot \frac{1}{6} + \cdots$$

$$= \frac{35}{12}$$

Statistics

Given data $X_1, \dots, X_n$, how do we learn something about the source of the data?

no assumptions = ill-posed

Ex. 1, 0, 1, 1, 1, 1, 0, 1 where did this come from?

A statistical model is a set of distributions

$$\mathcal{F} = \{ F_\theta : \theta \in \Theta \}, \quad , \Theta \subseteq \mathbb{R}^k \}$$

- parametric : $k < \infty$ $\qquad \theta = (\theta_1, \theta_2 \dots, \theta_k)$
- non-parametric : $k = \infty$

Ex $\theta = \{\theta(x) : x \in \mathbb{R}\}$ $\qquad F_\theta(x) = \theta(x)$

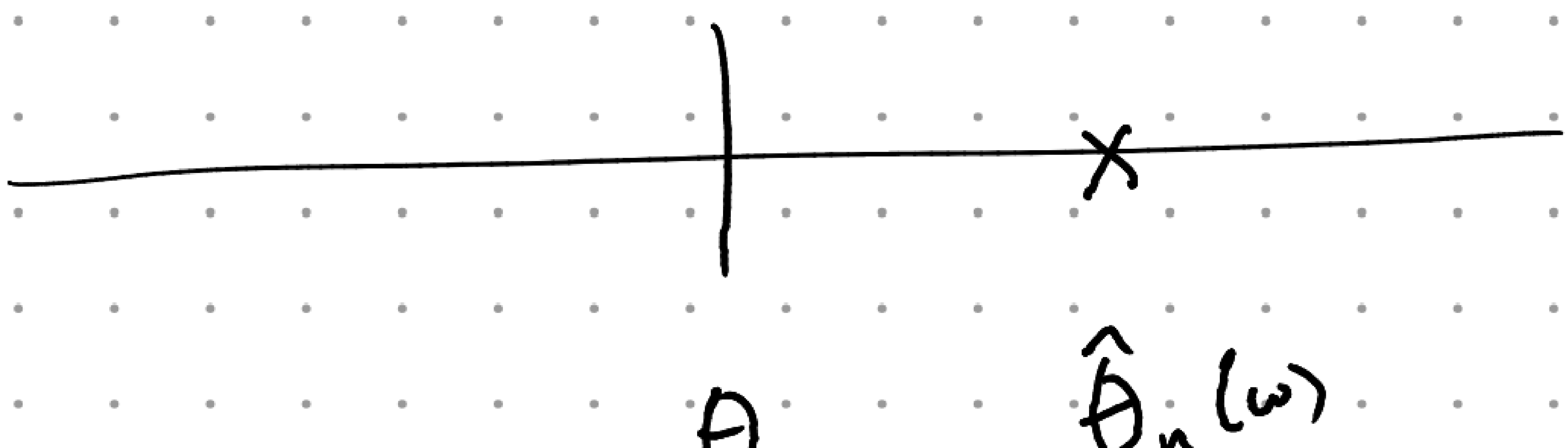
Assume $X_1, X_2, \dots, X_n \sim F_\theta$ (iid) for some $\theta \in \Theta$

How well can we learn something about θ / F_θ?

- Here how-well means statistically (i.e. if we repeated this many times)

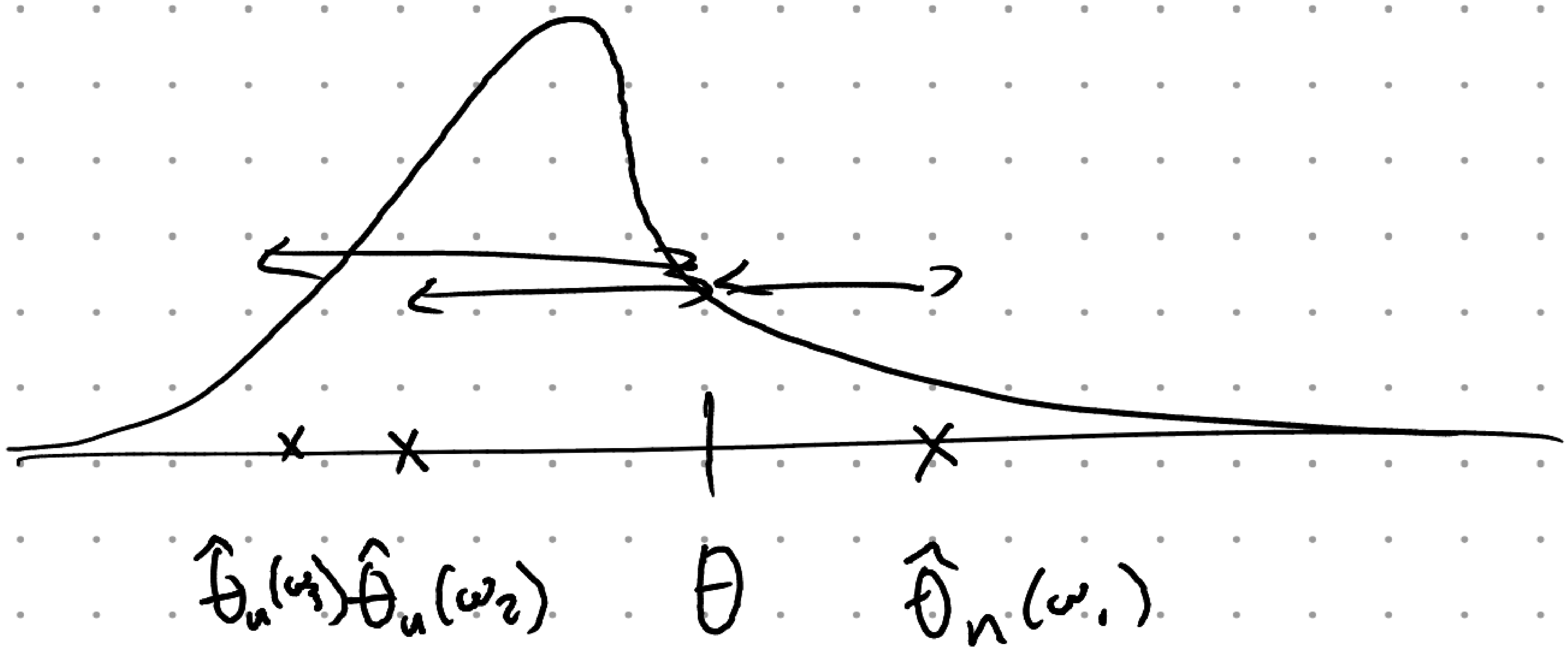
Suppose we use our data to get an estimate $\hat{\theta}_n$ for θ. How good is this estimate?

- Again, we can only think about this problem if $\hat{\theta}_n$ is a RV. For any fixed sample of our data this is not meaningful.



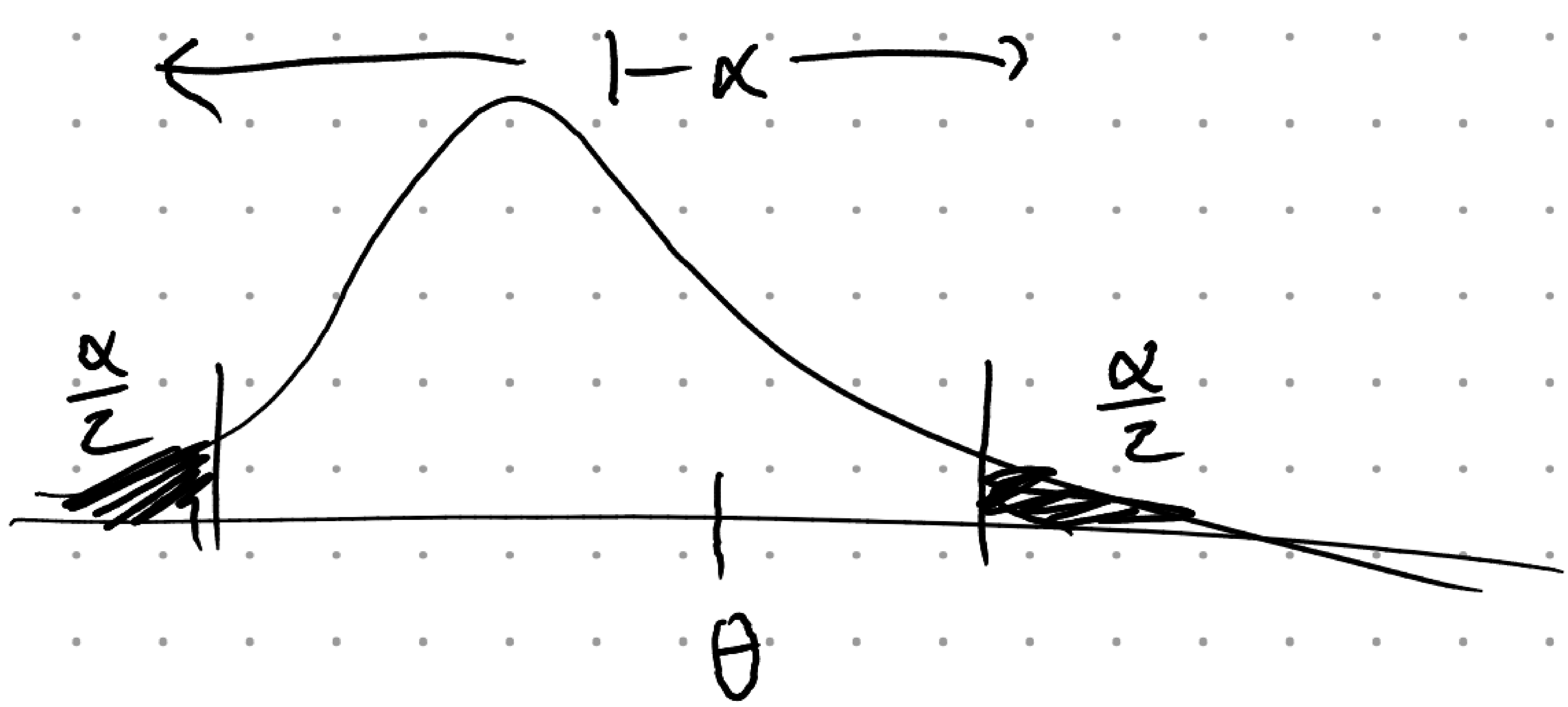
- is this good or bad?
- Did we get lucky?

As a RV, $\hat{\theta}_n$ has a distribution.



Is the distribution of $\hat{\theta}_n$ <u>usually</u> near θ?

- MSE : $\mathbb{E}\left[|\hat{\theta}_n - \theta|^2\right]$
- CI : $\mathbb{P}\left[\hat{\theta}_n \in (a_n, b_n)\right] \geq 1 - \varepsilon$



How do we come up with $\hat{\theta}_n$?

Parametric models:

- Method of moments
- MLE

General:

- Plug in estimator

Plug in estimator:

Something we can write our parameter θ as a functional of F.

<u>Ex</u>

$$\mu = \int x \, dF(x)$$

$$\sigma^2 = \int (X-\mu)^2 \, dF(x)$$

$$F(t) = \int \mathbb{1}[t \le x] \, dF(x) = \int_0^t 1 \, dF(x)$$

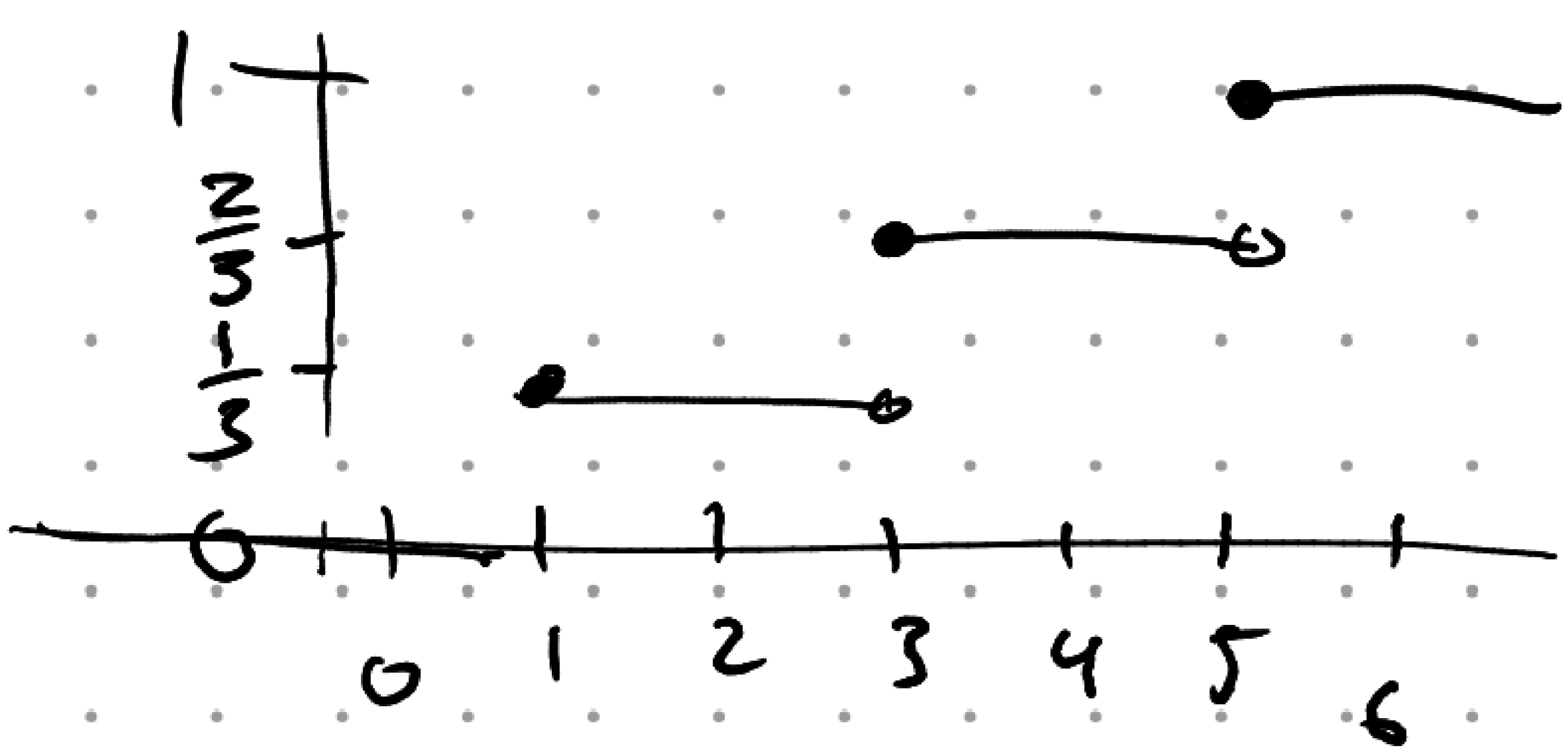
Get plug in estimator by replacing F with empirical CDF $\hat{F}_n = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}[X_i \le x]$

Conditional Expectation

Let X be a RV, A an event.

- conditional CDF: $F_A(x) = \mathbb{P}[X \le x \mid A] = \dfrac{\mathbb{P}[X \le x, A]}{\mathbb{P}[A]}$
- conditional expectation: $\mathbb{E}[X \mid A] = \int x \, dF_A(x)$

Ex. X output of dice, $A = \{\text{odd}\}$



Let X, Y RVs.

- conditional expectation $\mathbb{E}[X \mid Y]$ is a random variable which takes value $\mathbb{E}[X \mid Y = y]$ whenever $Y = y$.

Law of iterated expectation

$$\mathbb{E}[\mathbb{E}[X \mid Y]] = \mathbb{E}[X]$$

Total Expectation

$$\mathbb{E}[X] = \sum_{k=1}^{\infty} \mathbb{E}[X \mid A_k]\,\mathbb{P}[A_k]$$

if A_k countable partition of Ω

Ex. Chebyshev's inequality

Let $A_k = \{ |X-\mu| \geq \varepsilon \}$

$$\mathbb{V}[X] = \mathbb{E}[(X-\mu)^2]$$

$$= \mathbb{E}[(X-\mu)^2 \mid A_k]\,\mathbb{P}[A_k] + \mathbb{E}[(X-\mu)^2 \mid A_k^c]\,\mathbb{P}[A_k^c]$$

$$\geq \varepsilon^2 \cdot \mathbb{P}[|X-\mu| \geq \varepsilon] + 0 \cdot \mathbb{P}[A_k^c]$$

$$\Rightarrow \quad \mathbb{P}[|X-\mu| \geq \varepsilon] \leq \frac{\mathbb{V}[X]}{\varepsilon^2}$$