Lecture #1

- Let us recall the fundamentals of probability theory
- Let Ω be a sample space, and let $A \subseteq \Omega$ be an event. A probability measure on Ω is a function $\mathbb{P}$ that maps events to $[0,1]$ s.t. if Ω discrete

(1.) $$\sum_{w \in \Omega} \mathbb{P}(w) = 1$$

$= \mathbb{P}(\Omega)$] general

(2.) $\mathbb{P}(A) \geq 0 \quad \forall A \subseteq \Omega$

(3.) If $\{A_k\}_{k=1}^{\infty}$ are disjoint, then $\mathbb{P}\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} \mathbb{P}(A_k)$.

- Probabilities satisfy several algebraic properties, including:

(1.) Law of complement: if $A^c = \{w \in \Omega \mid w \notin A\}$, then $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$

(2.) Union bound: $\mathbb{P}\left(\bigcup_{k=1}^{\infty} A_k\right) \leq \sum_{k=1}^{\infty} \mathbb{P}(A_k)$, for any sets $\{A_k\}_{k=1}^{\infty}$.

- Given a notion of probability, we can consider random variables: a function $X : \Omega \to \mathbb{R}$. Note that $\mathbb{R}$ can be replaced with other spaces (e.g. $\mathbb{R}^d$, $\mathbb{C}$, ...)
- In what sense do we think of X as random? Saying $X : \Omega \to \mathbb{R}$ makes it seem just like a function on Ω. We need an associated

probability measure on X. There are a couple of fundamental ways to describe the random behavior of a r.v.

Defn: Let X be a r.v. on $(\Omega, \mathbb{P})$. The ~~[illegible]~~ cumulative distribution function of X is $F_X(x) = \mathbb{P}(X \leq x)$.

· Another, often more useful formulation of the probabilistic behavior of X is available, but depends on the structure (topology) of Ω.

Defn: Let Ω be discrete. The probability distribution function of a r.v. X on $(\Omega, \mathbb{P})$ is $f(x) = \mathbb{P}(X = x)$.

· If X is defined on Ω with uncountably many elements, then ~~[illegible]~~ generically $\mathbb{P}(X = x) = 0$, for x arbitrary. A somewhat more complicated description is needed. For simplicity, we will let $\Omega = \mathbb{R}$.

Defn: Let $\Omega = \mathbb{R}$. The probability density function of a r.v. X on $(\mathbb{R}, \mathbb{P})$ is $f: \mathbb{R} \to \mathbb{R}$ s.t.

$$\mathbb{P}(X \in [a,b]) = \int_a^b f(x)\,dx,$$

for any $a \leq b$.

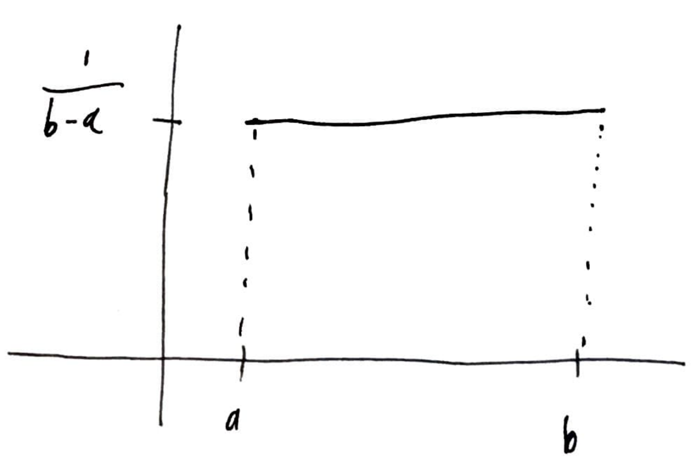
<u>Remark</u>: Despite the two different definitions, the idea is really the same for densities and distribution functions. This is established in the unified language of measure theory (see MATH 235-236).

<u>ex</u>: A very natural random object is one that gives all outputs equal probability. This is easy to do when Ω is finite. WLOG, let Ω finite be $\{1,2,3,\dots,n\}$ for some $n < \infty$. Then the uniform r.v. X on Ω is defined by the probability distribution function

$$\mathbb{P}(X=k) = \frac{1}{n} \quad \text{for all } k \in \{1,\dots,n\}.$$

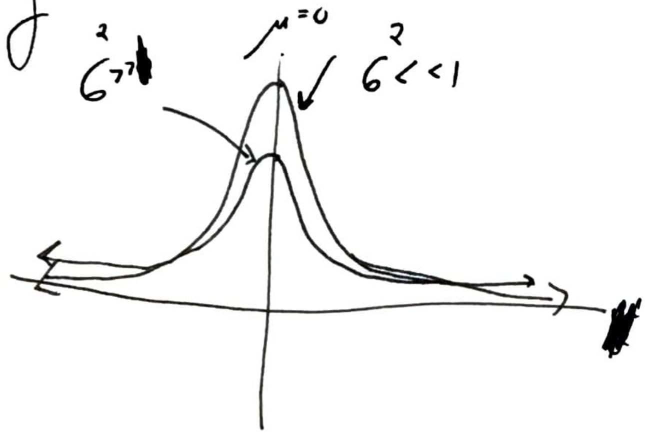
Of course, it is not possible to define a uniform distribution on $\mathbb{Z}_+ = \{1,2,\dots\}$. But, we can do a uniform distribution on any non-trivial interval $[a,b] \subseteq \mathbb{R}$, via the density function $\frac{1}{b-a} \cdot \mathbb{1}_{[a,b]}(x)$

$$= \begin{cases} \frac{1}{b-a}, & x \in [a,b] \\ 0, & \text{else} \end{cases}$$



The $\frac{1}{b-a}$ is a normalizing factor to ensure $\int_{\mathbb{R}} f(x)dx = 1$!

ex (Gaussian r.v.) Let $\mu \in \mathbb{R}$, $\sigma^2 \in \mathbb{R}_+ = (0,\infty)$ be fixed parameters. The Gaussian with parameters (μ, σ^2) is a $\mathbb{R}$-valued r.v. with density function $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left(-|x-\mu|^2/2\sigma^2\right)$.



- We will focus a (surprisingly?) large amount on Gaussians. Why? The Central Limit Theorem suggests many distributions can be usefully approximated by Gaussians. Why Gaussians? We can look at the proofs to get some insight (MATH 165).

- Sometimes we are interested in compressing a r.v. to just a few numbers.

<u>Defn</u>: Let X be a r.v. with probability distribution function $\mathbb{P}(X=k)$. The <u>expected value</u> and <u>variance</u> of X are, respectively,

- $\mathbb{E}(X) = \sum_k \mathbb{P}(X=k) \cdot k$.
- $\mathrm{Var}(X) = \sum_k \mathbb{P}(X=k) \cdot \left[k - \mathbb{E}(X)\right]^2$.

We may make similar definition if X has a density, with Σ replaced by $\int$:

Defn: Let X be a r.v. with density $f(x)$. The expected value and density are, respectively,

- $\mathbb{E}(X) = \int_{\mathbb{R}} x f(x)\, dx$

- $Var(X) = \int_{\mathbb{R}} \left(x - \mathbb{E}(X)\right)^2 f(x)\, dx$

ex: Let $X \sim N(\mu, \sigma^2)$, i.e. X is a Gaussian (also called a normal) r.v. with parameters $\mu \in \mathbb{R}$ and $\sigma^2 > 0$. Then

- $$\mathbb{E}(X) = \int_{-\infty}^{\infty} x f_X(x)\, dx$$
$$= \int_{-\infty}^{\infty} x \cdot \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{|x-\mu|^2}{2\sigma^2}\right) dx$$

do a u-sub...

$$\vdots$$
$$= \mu$$

- $$Var(X) = \int_{-\infty}^{\infty} (x-\mu)^2 \cdot \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left(-\frac{|x-\mu|^2}{2\sigma^2}\right) dx$$

IBP

$$\vdots$$
$$= \sigma^2$$

Remark: A nice alternative formulation for $Var(X)$ is given by expanding the quadratic $[x-\mathbb{E}(X)]^2 = x^2 - 2x\,\mathbb{E}(X) + \mathbb{E}(X)^2$:

$$Var(X) = \int_{\mathbb{R}} [x - \mathbb{E}(X)]^2 f_X(x)\,dx$$

$$= \int_{\mathbb{R}} \left[x^2 - 2x\,\mathbb{E}(X) + \mathbb{E}(X)^2\right] f_X(x)\,dx$$

$$= \int_{\mathbb{R}} x^2 f_X(x)\,dx - 2\mathbb{E}(X) \underbrace{\int_{\mathbb{R}} x f_X(x)\,dx}_{= \mathbb{E}(X)} + \mathbb{E}(X)^2 \underbrace{\int_{\mathbb{R}} f_X(x)\,dx}_{= 1}$$

$$= \int_{\mathbb{R}} x^2 f_X(x)\,dx - 2\mathbb{E}(X)\cdot\mathbb{E}(X) + \mathbb{E}(X)^2 \cdot 1$$

$$= \underbrace{\int_{\mathbb{R}} x^2 f_X(x)\,dx}_{= \mathbb{E}(X^2)} - \mathbb{E}(X)^2$$

So, we summarize: $Var(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$.