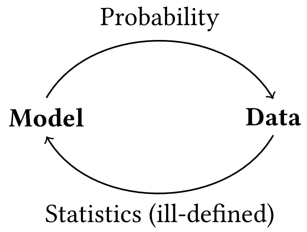
Probability Vs Statistics

Probability: Given Model, how likely is Data? $\rightarrow$ Well-formed since these are Mathematical questions.
Statistics: Given Data, how likely is Model? $\rightarrow$ Ill-formed since many Models can create the same data!



Probability Space

Probability Space: a triple (Ω, F, P) consisting of:

1. Ω the **Sample Space**
2. $F \subseteq 2^{\Omega}$ a σ**-algebra** on Ω i.e.
   (a) $\Omega \in F$
   (b) $E \in F \Rightarrow E^{\complement} \in F$
   (c) $E_1, E_2, \dots \in F \Rightarrow \bigcup_{i=1}^{\infty} E_i \in F$
3. P: $F \mapsto [0,1]$ a **Probability Measure** i.e.
   (a) $P(E) \geq 0$ for $E \in F$
   (b) $P(\Omega) = 1$
   (c) $P\left(\bigsqcup_{i=1}^{\infty} E_i\right) \Rightarrow \sum_{i=1}^{\infty} P(E_i)$ for $E_i \in F$

Given **Events** $E_i, E \in F$, P also satisfies:

1. **Upward and Downward continuity** of P:
   (a) $E_i \uparrow E \Rightarrow \lim_{n\to\infty} P(E_n) = P(E)$
   (b) $E_i \downarrow E \Rightarrow \lim_{n\to\infty} P(E_n) = P(E)$
2. **Monotonicity** of P:
   (a) $E_i \subseteq E_j \Rightarrow P(E_i) \leq P(E_j)$

Conditional Probability

We can compute Probabilities of Events Conditioned on other Events.
Conditional Probability of event A on event B with $P(B) > 0$ is:

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

A set of events $\{A_i\}$ are **Mutually Independent** if, for any n element subset of $\{A_i\}$:

$$P\left(\bigcap_{i=1}^{n} A_i\right) = \prod_{i=1}^{n} P(A_i)$$

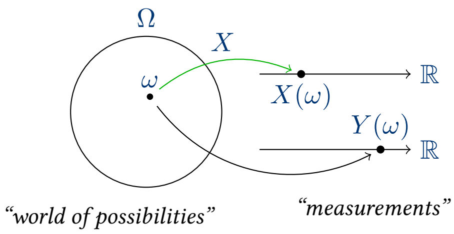
Law of Total Probability: Given Events A and **Partition** $\{B_i\}$ (i.e. where $\bigsqcup_{i=1}^{\infty} B_i = \Omega$)

$$P(A) = \sum_{i=1}^{\infty} P(A \mid B_i)P(B_i)$$

Random Variables

A **Random Variable** is a $\mathbb{B}$-Measurable function
X: $(\Omega, F) \mapsto (\mathbb{R}, \mathbb{B})$

1. For $A \in \mathbb{B}$ we can compute $P(X \in A)$
2. $P(X \in A) := P(X^{-1}(A)) = P(\{\omega \in \Omega : X(\omega) \in A\})$
3. $P(X^{-1}(\cdot)) := P_X(\cdot)$ which is called the **Push-Forward Measure** of P by X on $\mathbb{R}$
4. Hence X induces a new Probability Space $(\mathbb{R}, \mathbb{B}, P_X)$ from the original (Ω, F, P)



Random Variables can be uniquely determined by their **CDF**:
$F_X(t) := P_X((-\infty, t]) = P(X \leq t)$

1. Right Continuous
2. Non-Negative
3. $\lim_{t\to\infty} F_X(t) = 1$, $\lim_{t\to-\infty} F_X(t) = 0$
4. Any function satisfying above properties is the CDF for some random variable.

Random Variables studied are usually either **Continuous** or **Discrete** (although they can be neither.)

1. If F_X **Absolutely Continuous** then X is Continuous.
   (a) **Absolutely Continuous**: F differentiable a.e. and $\exists f(x)$ s.t. $F_X(x) = \int_{-\infty}^{x} f(u)du$
   (b) f is called the **PDF**
   (c) Where f is continuous, $\frac{d}{dx}F_X(x) = f(x)$
   (d) hence, f is unique a.e. (may not be everywhere!)
   (e) If X also Non-Negative then **Hazard** of X is $\lambda(t) = \frac{f(t)}{1-F(t)}$
      i. $1 - F(t) = \exp\left(-\int_0^t \lambda(x)\,dx\right)$
      ii. $\lambda(t)$ interpreted as instantaneous survival rate at time t.
      iii. $\lambda(t) = c\;\forall t \iff X \sim Exp(c)$
2. If $Range(X)$ is countable then X is Discrete.
   (a) $f(x) := P(\{X = x\})$

Random Vectors

Joint CDF for $\vec{X} = (X_1, X_2, ..., X_n)$ is
$F(t_1, ..., t_n) = P(X_1 \leq t_1, ..., X_n \leq t_n)$

1. **Marginal PDF** of $\vec{X}_{1:p} = (X_1, \cdots X_p)$ is

$$f_{\vec{X}_{1:p}}(\vec{u}_{1:p}) = \int_{\vec{X}_{(p+1):n}} f_{\vec{X}}\left(\vec{u}_{1:p}, \vec{X}_{(p+1):n}\right) d\vec{X}_{(p+1):n}$$

2. **Conditional PDF** on $\vec{X}_{1:p}$ given $\vec{X}_{(p+1):n}$ is

$$f_{\vec{X}_{1:p}|\vec{X}_{(p+1):n}}\left(\vec{u}_{1:p}, \vec{u}_{(p+1):n}\right) = \frac{f_{\vec{X}}\left(\vec{u}_{1:p}, \vec{u}_{(p+1):n}\right)}{f_{\vec{X}_{(p+1):n}}\left(\vec{u}_{(p+1):n}\right)}$$

3. **kth Order Statistic** $X_{(k)}$ of $\vec{X}$ is the kth smallest value
   (a) $f_{X_{(1)}}(u) = \sum_{i=1}^{n} f_{X_i}(u) \prod_{j\neq i}(1 - F_{X_j}(u))$
   (b) $f_{X_{(n)}}(u) = \sum_{i=1}^{n} f_{X_i}(u) \prod_{j\neq i} F_{X_j}(u)$

Moments of a Random Variable

$\mathbb{E}_X(X^r) := \mathbb{E}(X^r)$ is the r**th Moment** of X under the distribution of X

1. $\mathbb{E}(X) = \int_0^\infty 1 - F_X(t)\, dt - \int_{-\infty}^0 F_X(t)\, dt$
2. If X Continuous, $\mathbb{E}(X) = \int_{-\infty}^{\infty} t \cdot f_X(t)\, dt$
3. **LOTUS**: $\mathbb{E}(g(X)) = \int_{-\infty}^{\infty} g(t) \cdot f_X(t)\, dt$

Can generate Moments using the **MGF** of X:
$M_X(t) = \mathbb{E}\left(\exp(Xt)\right)$, if $\exists \epsilon > 0$ s.t. $\forall |t| < \epsilon$, $M_X(t) < \infty$

1. $\exists \epsilon > 0$ s.t. $\forall |t| < \epsilon$, $M_X(t) = M_Y(t) \Rightarrow X$ and Y have same distribution
2. $\mathbb{E}(|X^r|) = \frac{\partial^r}{\partial^r t} M_X(t)\big|_{t=0}$, if M_X exists.
3. If $\{X_i\}$ independent RVs, then $M_{\sum X_i}(t) = \prod M_{X_i}(t)$
4. Moments most commonly analyzed are:
   (a) **Mean** of X: $\mathbb{E}(X) := \mu_X$
   (b) **Variance** of X: $Var(X) = \mathbb{E}((X - \mu_X)^2) = \sigma_X^2$

If X and Y are Random Variables on the same Probability Space

1. **Law of Total Expectation**: If $\mathbb{E}(|X|) < \infty$
   $\mathbb{E}(X) = \mathbb{E}_Y(\mathbb{E}_{X|Y}(X \mid Y))$
2. **Law of Total Variance**: If $Var(X) < \infty$
   $Var(X) = \mathbb{E}(Var(X \mid Y)) + Var(\mathbb{E}(X \mid Y))$

For random vectors we have

Notes

1. Continuity of X as a function has nothing to do with its continuity as a Random Variable (which depends on the absolute continuity of its CDF) [1]
2. $\mathbb{E}(|X|) < \infty$ iff $\mathbb{E}(X)$ exists, by the definition of lebesgue integrability and the measurability of X

References

[1] pidgeot (https://math.stackexchange.com/users/181948/pidgeot), "On clarifying the relationship between distribution functions in measure theory and probability theory." Mathematics Stack Exchange. URL:https://math.stackexchange.com/q/976739 (version: 2014-10-16).