Probability and Random Processes

Irvin Avalos

Chapter 1

Probability Theory

Defintion 1.0.1 (Sample Space, Ω)**.** *The set of all possible outcomes of an experiment.*

1.1 Discrete Random Variables

- A <u>discrete random variable</u> associates a real number with each outcome in Ω, e.g., $X = i$ is a random variable if the throw of a die is i for $i = 1, 2, \ldots, 6$.
- Note that X^2 is still a random variable and in general, a function of a random variable is also a random variable.
- <u>Probability mass function</u> is the set $\{a, P_X(a)\}$ where $a \in \mathcal{A}$ and $\mathcal{A}$ is the set of all possible values taken by X.

Example 1.1.1. *Chess match between Alice and Bob. The first to win a game wins the match and conversely, the match is drawn if there are* 10 *consecutive draws.*

$$P(\textit{Alice wins a game}) = 0.3 \qquad P(\textit{Bob wins a game}) = 0.4 \qquad P(\textit{Draw}) = 0.3$$

1. *What is the PMF of the duration of a match?*

   Let L denote the duration of the match. Suppose $L = 10$*, then*

   $$P_L(10) = (0.3)^9.$$

   Otherwise, if $L = k < 10$*,*

   $$P_L(k) = (0.3)^{k-1}(0.7) \quad \textit{for} \;\; k = 1, 2, \ldots, 9.$$

2. *Calculate* $P(\textit{Alice wins the match})$*.*

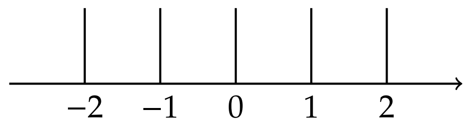
$$\begin{aligned} P(\textit{Alice wins the match}) &= (0.3) + (0.3)(0.3) + \ldots + (0.3)^9(0.3) \\ &= \sum_{k=0}^{9} (0.3)^k (0.3) \end{aligned}$$

3. *Assuming no limit to the number of games played, find P(Alice wins the match) again.*

   (a) $P(\textit{Alice wins the match}) = \sum_{k=0}^{\infty}(0.3)^k(0.3) = (0.3)\sum_{k=0}^{\infty}(0.3)^k = \frac{0.3}{1-0.7} = \frac{3}{7}$

- <u>Functions of a RV:</u>

  Let $P_X(x) = 0.2 \quad x = -2, -1, 0, 1, 2.$



Chapter 2

Information Theory

Defintion 2.0.1 (Entropy)**.**

$$H(X) = \mathrm{E}\left[-\log_2 p(x)\right] = \sum_{x \in X} -p(x)\log_2 p(x)$$

If $X \sim B(p)$, then

$$H(X) = p - \log_2 p + (1-p)\log_2(1-p) \triangleq h(p)$$

*called the **binary entropy function**.*

Defintion 2.0.2 (Joint Entropy)**.**

$$H(X,Y) = \mathrm{E}\left[-\log_2 p(x,y)\right] = \sum_x \sum_y -p_{x,y}(x,y)\log_2 p(x,y)$$

If X, Y are independent, then $H(X,Y) = H(X) + H(Y)$.

Proof.

$$\begin{aligned} H(X,Y) &= \mathrm{E}\left[-\log_2 p(x,y)\right] \\ &= \mathrm{E}\left[-\log_2 p(x)\, p(y)\right] \\ &= \mathrm{E}\left[-\log_2 p(x) - \log_2 p(y)\right] \\ &= \mathrm{E}\left[-\log_2 p(x)\right] + \mathrm{E}\left[-\log_2 p(y)\right] \\ &= H(X) + H(Y) \end{aligned}$$

□

Defintion 2.0.3 (Conditional Entropy)**.**

$$H(Y \mid X) = \mathrm{E}_{XY}\left[-\log_2 p(y \mid x)\right] H(X,Y) - H(X)$$

Defintion 2.0.4 (Mutual Information)**.**

$$\begin{aligned} I(X;Y) &\triangleq H(X) - H(X \mid Y) \\ &= H(Y) - H(Y \mid X) \\ &= H(X) + H(Y) - H(X,Y) \end{aligned}$$

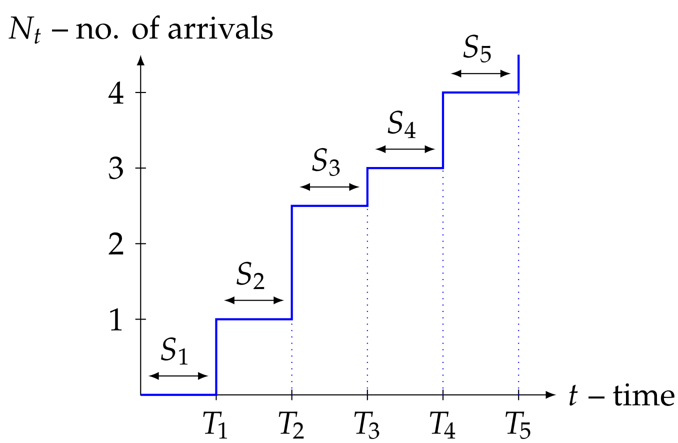
Theorem 2.0.5 (Asymptotic Equipartition)**.** *If $X_1, X_2, \ldots, X_n$ are iid $\sim p(X)$, then*

$$\frac{-\log_2 p(X_1, X_2, \ldots, X_n)}{n} \xrightarrow{p} H(X).$$

Chapter 3

Poisson Proceess

A Poisson process is the continuous time analog of "coin flipping" (or Bernoulli) processes. This makes it a good model for arrival processes: photons hitting a detector, packets in a network, number of emails per hour, etc.



Each T_i for $i = 1, 2, 3, 4, 5$ represents an arrival and generally, each arrival time is defined as

$$T_n = \sum_{i=1}^{n} S_i$$

where the interarrival times $S_1, S_2, \ldots, S_n \overset{\text{iid}}{\sim} Exponential(\lambda)$. Thus, every S_i has the probability density function

$$f_{S_i}(t) = \lambda e^{\lambda t};\ t > 0; \quad i = 1, 2, 3, \ldots$$

and cumulative distribution function

$$F_{S_i}(t) = 1 - e^{-\lambda t}; \quad i = 1, 2, 3, \ldots.$$

Defintion 3.0.1 (Number of Arrivals)**.**

$$N_t = \begin{cases} \max_{n \geq 1}\{n \mid T_n \leq t\} & t \geq 0 \\ 0 & t < T_1 \end{cases}$$

- Recall: $Exponential(\lambda)$ is a memoryless RV

  1. $F_\tau(t) = \begin{cases} 1 - e^{-\lambda t} & t \geq 0 \\ 0 & t < 0 \end{cases} \Rightarrow f_\tau(t) = \lambda e^{-\lambda t}$
  2. $\mathrm{E}[\tau] = \frac{1}{\lambda}$ and $\mathrm{Var}(\tau) = \frac{1}{\lambda^2}$
  3. $\mathrm{P}(\tau > t + s \mid \tau > s) = \mathrm{P}(\tau > t)$
  4. $\mathrm{P}(\tau \leq t + \epsilon \mid \tau > t) = \lambda\epsilon + o(\epsilon)$; $\lim_{\epsilon \to 0} \frac{o(\epsilon)}{\epsilon} = 0$

Proof.

$$\begin{aligned} \mathrm{P}(\tau > t + \epsilon \mid \tau > t) &= \mathrm{P}(\tau > \epsilon) \\ &= e^{-\lambda\epsilon} \\ &= 1 - \lambda\epsilon + o(\epsilon) \\ \therefore \mathrm{P}(\tau \leq t + \epsilon \mid \tau > t) &= 1 - \mathrm{P}(\tau > t + \epsilon \mid \tau > t) \\ \lambda\epsilon + o(\epsilon) \end{aligned}$$

□

- $\mathrm{P}(1 \text{ arrival}) = \lambda\epsilon + o(\epsilon)$
- $\mathrm{P}(2 \text{ arrivals}) = 1 - \lambda\epsilon + o(\epsilon)$
- $\mathrm{P}(3 \text{ arrivals}) = o(\epsilon)$

We can only have one arrival in a particular subinterval with high probability $\lambda\epsilon + o(\epsilon)$. On the other hand, there are zero arrivals with probability $1 - \lambda\epsilon + o(\epsilon)$.

This means that a Poisson process has independent and stationary increments, i.e., for any $0 \leq t_1 \leq t_2 \leq \ldots \leq t_n \leq \ldots$; if you look at the number of arrivals in t_{n+1} and t_n, $\{N_{t_{n+1}} - N_{t_n}\}$ are independent and the distribution depends only on the size of the interval $(t_{n+1} - t_n)$.

Theorem 3.0.2. *Let $N_t :=$ number of arrivals in $(0, t)$, then*

$$P(N_t = k) = e^{-\lambda t}\frac{(\lambda t)^k}{k!}; \quad k = 0, 1, 2, \ldots$$

Proof.

$$P(N_t = k) = \int_{t_1}\int_{t_2} \ldots \int_{t_k} f_{T_1, T_2, \ldots, T_k \mid N_t = k}(t_1, t_2, \ldots, t_k \mid N_t = k) dt_1\, dt_2\, \ldots\, dt_k$$

We constrain the region of integration by letting $S = 0 \le t_1 \le t_2 \le \ldots \le t_k \le t \le t_{k+1}$. Shifting focus to the k^{th} dimensional pdf,

$$f_{T_1,\ldots,T_k|N_t=k}(t_1,\ldots,t_k \mid N_t = k) = \mathrm{P}(T_1 \in (t_1, t_1 + dt_1), \ldots, T_k \in (t_k, t_k + dt_k), T_{k+1} > t)$$

where we leverage the fact that $\mathrm{P}(T_{k+1} \in (t_{k+1}, t)) = \mathrm{P}(T_{k+1} > t)$

$$\begin{aligned}
&\Rightarrow \mathrm{P}(T_1 \in (t_1, t_1 + dt_1)) \times \ldots \times \mathrm{P}(T_k \in (t_k, t_k + dt_k)) \times \mathrm{P}(T_{k+1} > t) \\
&= \mathrm{P}(S_1 \in (t_1, t_1 + dt_1)) \times \ldots \times \mathrm{P}(S_k \in (t_k - t_{k-1}, t_k - t_{k-1} + dt_k)) \times \mathrm{P}(S_{k+1} > t - t_k) \\
&= (\lambda e^{-\lambda t_1} dt_1) \ \ldots \ (\lambda e^{-\lambda(t_k - t_{k-1})} dt_k) \ e^{-\lambda(t-t_k)} \\
&= \lambda^k e^{-\lambda t}
\end{aligned}$$

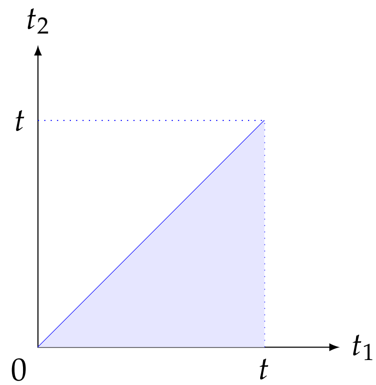
The result $\lambda^k e^{-\lambda t}$ is a constant to the k^{th} dimensional integral, so barring any constraints on $t_1, t_2, \ldots, t_k$, Volume $(S) = t^k$. But, we must respect the fact that $0 \le t_1 \le t_2 \le \ldots \le t_k$.

By symmetry, every permutation of $(t_1, t_2, \ldots, t_k)$ has equal volume

$$\text{Volume } (S) = \frac{t^k}{k!} \Rightarrow \mathrm{P}(N_t = k) = \frac{(\lambda t)^k e^{-\lambda t}}{k!}; \quad k = 0, 1, 2, \ldots$$

□

- *Conditioned on the number of arrivals being k in an interval, the density is uniform*
- *Intuition behind symmetry argument:* $k = 2$



From the image above we know that $\mathrm{P}(t_1 \le t_2) + \mathrm{P}(t_2 < t_1) = 1$. *Thus,* $\mathrm{P}(t_1 < t_2) = \frac{1}{2} = \frac{1}{2!}$ *which generalizes to k dimensions.*

Defintion 3.0.3 (Poisson Merging). *Merging two or more independent Poisson processes yields* $PP(\lambda_1 + \lambda_2 + \ldots + \lambda_n)$

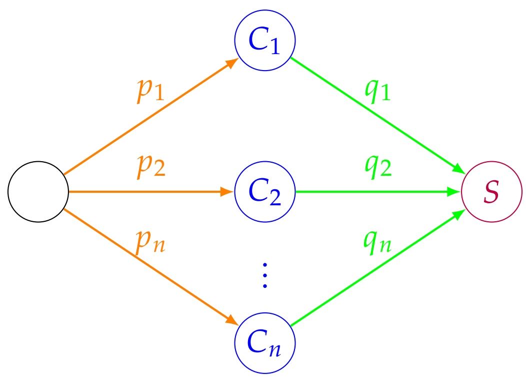
Defintion 3.0.4 (Poisson Splitting). *Let* $N \sim PP(\lambda)$ *such that* N *is split into* N_1 *and* N_2. *For each arrival we flip a coin independently with probability* $H = p$. *If the coin lands on heads, send the arrival to the* N_1 *queue otherwise it is sent to the* N_2 *queue.*

Chapter 4

Statistical Inference

4.1 Detection & Bayes' Theorem

Consider N possible exclusive causes of a particular health symptom.



- Each cause i has a prior probability p_i and it has a probability q_i of causing the observed symptom.
- We want to esimate the posterior probability π_i of cause i given the symptom S,

$$\pi_i = \mathrm{P}(C_i \mid S) = \frac{\mathrm{P}(C_i \cap S)}{\mathrm{P}(S)} = \frac{\mathrm{P}(S \mid C_i)\ \mathrm{P}(C_i)}{\sum_j \mathrm{P}\left(S \mid C_j\right)\ P\left(C_j\right)}.$$

From the diagram above the posterior distribution for cause i can be simplified to

$$\pi_i = \frac{q_i\, p_i}{\sum_j q_j\, p_j}. \tag{4.1}$$

4.2 MAP & MLE

Defintion 4.2.1 (Maximum a posteriori (MAP))**.**

$$\mathrm{MAP} \triangleq \arg\max_i \ \pi_i = \arg\max_i \ p_i \, q_i$$

Defintion 4.2.2 (Maximum Likelihood Estimation (MLE))**.**

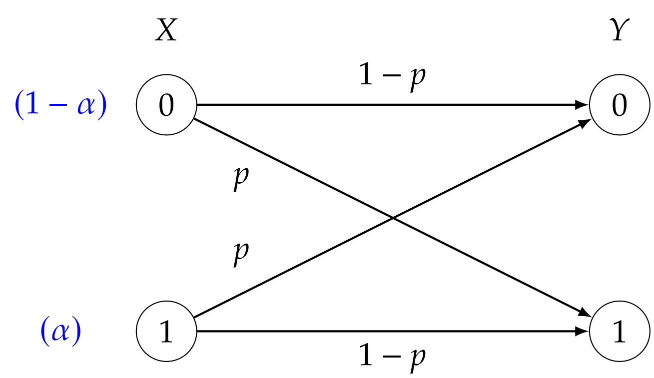
$$\mathrm{MLE} \triangleq \arg\max_i q_i \equiv \mathrm{MAP} \ \textit{assuming uniform priors} \ \ p_i = \frac{1}{N} \ \forall i$$

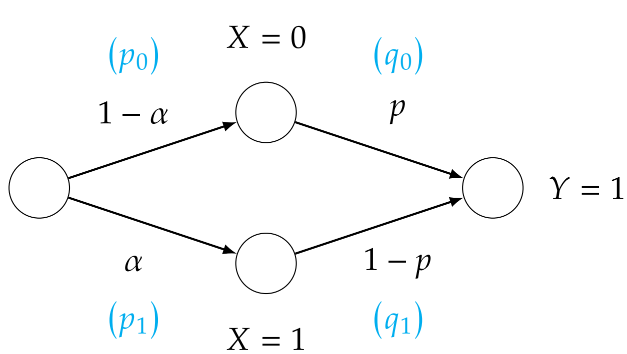
More generally, if X, Y are discrete RVs then

$$\mathrm{MAP}\left[X \mid Y = y\right] = \arg\max_x \ \mathrm{P}\left(X = x \mid Y = y\right)$$
$$\mathrm{MLE}\left[X \mid Y = y\right] = \arg\max_x \ \mathrm{P}\left(Y = y \mid X = x\right)$$

- Called detection because everything is discrete, i.e., detection and classification are synonymous.
- MAP: which cause best explaisn the observed symptom.
- MLE: which cause best generates/induces the observed symptom.

Example 4.2.1 (MAP/MLE Analysis of BSC)**.** *Consider a* BSC (p) *with* $p < \frac{1}{2}$.





$$\Rightarrow \mathrm{MAP}\left[X \mid Y=1\right] = \underset{i\in\{0,1\}}{\arg\max}\ p_i q_i.$$

This yields the inequality statement,

$$p_0 q_0 \underset{\hat{X}_1=1}{\overset{\hat{X}_1=0}{\gtrless}} p_1 q_1$$

where $\hat{X}_1$ is the MAP estimate of X when $Y=1$ is 0 or 1. Thus

$$(1-\alpha)\, p \underset{1}{\overset{0}{\gtrless}} \alpha\left(1-p\right)$$