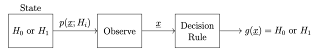
Intro: **Random Experiment:** An outcome for each run.
Sample Space Ω: Set of all possible outcomes.
Event: Subsets of Ω.
Prob. of Event A: $P(A) = \frac{\text{Number of outcomes in A}}{\text{Number of outcomes in } \Omega}$
Axioms: $P(A) \geq 0 \; \forall A \in \Omega$, $P(\Omega) = 1$,
If $A \cap B = \emptyset$, then $P(A \cup B) = P(A) + P(B) \; \forall A, B \in \Omega$
Cond. Prob. $P(A|B) = \frac{P(A \cap B)}{P(B)}$
$* \; P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$
Independence: $P(A|B) = P(A) \Leftrightarrow P(A \cap B) = P(A)P(B)$
Total Prob. Thm: If $H_1, H_2, \ldots, H_n$ form a partition of Ω, then $P(A) = \sum_{i=1}^{n} P(A|H_i)P(H_i)$.
Bayes' Rule: $P(H_k|A) = \frac{P(H_k \cap A)}{P(A)} = \frac{P(A|H_k)P(H_k)}{\sum_{i=1}^{n} P(A|H_i)P(H_i)}$
*Posteriori: $P(H_k|A)$, Likelihood: $P(A|H_k)$, Prior: $P(H_k)$
1 RV: **CDF:** $F_X(x) = P[X \leq x]$
PMF: $P_X(x_j) = P[X = x_j] \; j = 1, 2, \ldots$
PDF: $f_X(x) = \frac{d}{dx} F_X(x)$
*$P[a \leq X \leq b] = \int_a^b f_X(x)\, dx$ IS THIS CORRECT?
Cond. PMF: $P_X(x|A) = P[X = x|A] = \frac{P[X=x, A]}{P[A]}$ IS THIS CORRECT?
Cond. PDF: $f_X(x|A) = \frac{f_{X,A}(x,a)}{f_A(a)}$ IS THIS CORRECT?
Exp.: $E[h(X)] = \int_{-\infty}^{\infty} h(x) f_X(x) dx \mid \sum_{k=-\infty}^{\infty} h(k) P_X(x_i = k)$
Variance: $\sigma_X^2 = \text{Var}[X] = E[(X - E[X])^2] = E[X^2] - E[X]^2$
Cond. Exp.: $E[X|A] = \int_{-\infty}^{\infty} x f_X(x|A)\, dx$
2 RVs: **Joint PMF:** $P_{X,Y}(x,y) = P[X = x, Y = y]$
Joint PDF: $f_{X,Y}(x,y) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x,y)$
*$P[(X,Y) \in A] = \int \int_{(x,y) \in A} f_{X,Y}(x,y)\, dx\, dy$
Exp.: $E[g(X,Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y)\, dx\, dy$
Correlation (Corr.): $E[XY]$
Covar.: $\text{Cov}[X,Y] = E[(X - \mu_X)(Y - \mu_Y)] = E[XY] - E[X]E[Y]$
Corr. Coeff.: $\rho_{X,Y} = E\left[\left(\frac{X - \mu_X}{\sigma_X}\right)\left(\frac{Y - \mu_Y}{\sigma_Y}\right)\right] = \frac{\text{Cov}[X,Y]}{\sigma_X \sigma_Y}$
Marginal PMF: $P_X(x) = \sum_{j=1}^{\infty} P_{X,Y}(x, y_j)$
Marginal PDF: $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dy$
Cond. PMF: $P_{X|Y}(x|Y) = P[X = x|Y = y] = \frac{P_{X,Y}(x,y)}{P_Y(y)}$
Cond. PDF: $f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$
Bayes' Rule
$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \frac{f_{X|Y}(x|y) f_Y(y)}{\int_{-\infty}^{\infty} f_{X|Y}(x|y') f_Y(y')\, dy'}$
*$P_{Y|X}(y|x) = \frac{P_{X,Y}(x,y)}{P_X(x)} = \frac{P_{X|Y}(x|y) P_Y(y)}{\sum_{j=1}^{\infty} P_{X|Y}(x|y_j) P_Y(y_j)}$
Ind.: $f_{X|Y}(x|y) = f_X(x) \; \forall y \Leftrightarrow f_{X,Y}(x,y) = f_X(x) f_Y(y)$
*If independent, then uncorrelated.
Uncorrelated: $\text{Cov}[X,Y] = 0 \Leftrightarrow \rho_{X,Y} = 0$
Orthogonal: $E[XY] = 0$
Cond. Exp.: $E[Y] = E[E[Y|X]]$ or $E[E[h(Y)|X]]$
*$E[E[Y|X]]$ w.r.t. X | $E[Y|X]$ w.r.t. Y.
Estimation: Estimate unknown parameter θ from n i.i.d. measurements $X_1, X_2, \ldots, X_n$, $\hat{\Theta}(\underline{X}) = g(X_1, X_2, \ldots, X_n)$
Estimation Error: $\hat{\Theta}(\underline{X}) - \theta$.
Unbiased: $\hat{\Theta}(\underline{X})$ is unbiased if $E[\hat{\Theta}(\underline{X})] = \theta$.
***Asymptotically unbiased:** $\lim_{n \to \infty} E[\hat{\Theta}(\underline{X})] = \theta$.
Consistent: $\hat{\Theta}(\underline{X})$ is consistent if $\hat{\Theta}(\underline{X}) \to \theta$ as $n \to \infty$ or $\forall \epsilon > 0, \lim_{n \to \infty} P[|\hat{\Theta}(\underline{X}) - \theta| < \epsilon] \to 1$.
Sample Mean: $M_n = \frac{1}{n} S_n = \frac{1}{n} \sum_{i=1}^{n} X_i$.
*Given a sequence of i.i.d. RVs, $X_1, X_2, \ldots, X_n$, M_n is unbiased and consistent.
Chebychev's Inequality: $P[|X - E[X]| \geq \epsilon] \leq \frac{\text{Var}[X]}{\epsilon^2}$
Weak Law of Large #s: $\lim_{n \to \infty} P[|M_n - \mu| < \epsilon] = 1 \; \forall \epsilon > 0$.
ML Estimation: Choose parameter θ that is most likely to generate the obs. $x_1, x_2, \ldots, x_n$.
*Disc: $\hat{\Theta} = \arg\max_\theta P_{\underline{X}}(\underline{x}|\theta) \xrightarrow{\log} \hat{\theta} = \arg\max_\theta \sum_{i=1}^{n} \log P_X(x_i|\theta)$
*Cont: $\hat{\Theta} = \arg\max_\theta f_{\underline{X}}(\underline{x}|\theta) \xrightarrow{\log} \hat{\theta} = \arg\max_\theta \sum_{i=1}^{n} \log f_X(x_i|\theta)$
Maximum A Posteriori (MAP) Estimation:
*Disc: $\hat{\theta} = \arg\max_\theta P_{\Theta|\underline{X}}(\theta|\underline{x}) = \arg\max_\theta P_{\underline{X}|\Theta}(\underline{x}|\theta) P_\Theta(\theta)$
*Cont: $\hat{\theta} = \arg\max_\theta f_{\Theta|\underline{X}}(\theta|\underline{x}) = \arg\max_\theta f_{\underline{X}|\Theta}(\underline{x}|\theta) f_\Theta(\theta)$
*$f_{\Theta|\underline{X}}(\theta|\underline{x})$: Posteriori, $f_{\underline{X}|\Theta}(\underline{x}|\theta)$: Likelihood, $f_\Theta(\theta)$: Prior
Bayes' Rule: $P_{\Theta|\underline{X}}(\theta|\underline{x}) = \begin{cases} \frac{P_{\underline{X}|\Theta}(\underline{x}|\theta) P_\Theta(\theta)}{P_{\underline{X}}(\underline{x})} & \text{if } \underline{X} \text{ disc.} \\ \frac{f_{\underline{X}|\Theta}(\underline{x}|\theta) P_\Theta(\theta)}{f_{\underline{X}}(\underline{x})} & \text{if } \underline{X} \text{ cont.} \end{cases}$
$f_{\Theta|\underline{X}}(\theta|\underline{x}) = \begin{cases} \frac{P_{\underline{X}|\Theta}(\underline{x}|\theta) f_\Theta(\theta)}{P_{\underline{X}}(\underline{x})} & \text{if } \underline{X} \text{ disc.} \\ \frac{f_{\underline{X}|\Theta}(\underline{x}|\theta) f_\Theta(\theta)}{f_{\underline{X}}(\underline{x})} & \text{if } \underline{X} \text{ cont.} \end{cases}$
*Independent of θ: $f_{\underline{X}}(\underline{x}) = \int_{-\infty}^{\infty} f_{\underline{X}|\Theta}(\underline{x}|\theta) f_\Theta(\theta)\, d\theta$

Beta Prior Θ is a Beta R.V. w/ $\alpha, \beta > 0$
$f_\Theta(\theta) = \begin{cases} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1}(1-\theta)^{\beta-1} & \text{if } 0 < \theta < 1 \\ 0 & \text{otherwise} \end{cases}$
*$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t}\, dt$
Prop.: 1. $\Gamma(x+1) = x\Gamma(x)$. For $m \in \mathbb{Z}^+$, $\Gamma(m+1) = m!$.
2. $\beta(\alpha, \beta) = \frac{(\alpha+\beta-1)!}{(\alpha-1)!(\beta-1)!} = \beta\binom{\alpha+\beta-1}{\alpha-1}$
3. Expected Value: $E[\Theta] = \frac{\alpha}{\alpha+\beta}$ for $\alpha, \beta > 0$
4. Mode (max of PDF): $\frac{\alpha-1}{\alpha+\beta-2}$ for $\alpha, \beta > 1$

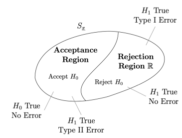
Drawing Beta Dist. 1. Label x-axis from 0 to 1. 2. Identify mode.
3. Determine shape based on α and β: $\alpha = \beta = 1$ (uniform), $\alpha = \beta > 1$ (bell-shaped, peak at 0.5), $\alpha = \beta < 1$ (U-shaped w/ high density near 0 and 1), $\alpha > \beta$ (left-skewed), $\alpha < \beta$ (right-skewed).
Least Mean Squares (LMS) Estimation: Assume prior $P_\Theta(\theta)$ or $f_\Theta(\theta)$ w/ obs. $\underline{X} = \underline{x}$.
*$\hat{\theta} = g(\underline{x}) = \mathbb{E}[\Theta|\underline{X} = \underline{x}]$ | $\hat{\Theta} = g(\underline{X}) = \mathbb{E}[\Theta|\underline{X}]$

Uniform PDF $f_X(x) = \begin{cases} \frac{1}{b-a} & \text{if } a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$
*$E[X] = \frac{a+b}{2}$, $\text{Var}[X] = \frac{(b-a)^2}{12}$
Conditional Exp. $E[X|Y] = \int_{-\infty}^{\infty} x f_{X|Y}(x|y)\, dx$

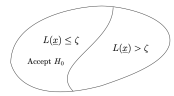
Binary Hyp. Testing: H_0: Null Hyp., H_1: Alt. Hyp.



TI Err. (False Rejection): Reject H_0 when H_0 is true.
*$\alpha(R) = P[X \in R \mid H_0]$
TII Err. (False Accept.): Accept H_0 when H_1 is true.
*$\beta(R) = P[X \in R^c \mid H_1]$



Likelihood Ratio Test: For each value of $\underline{x}$,
*$L(\underline{x}) = \frac{P_{\underline{X}|H_1}(\underline{x}|H_1)}{P_{\underline{X}|H_0}(\underline{x}|H_0)}$ | $L(\underline{x}) = \frac{f_{\underline{X}|H_1}(\underline{x}|H_1)}{f_{\underline{X}|H_0}(\underline{x}|H_0)}$
***ML Rule:** $L(\underline{x}) \leq 1 \Rightarrow$ Accept H_0 | $L(\underline{x}) > 1 \Rightarrow$ Reject H_0
***General:** $L(\underline{x}) \leq \xi \Rightarrow$ Accept H_0 | $L(\underline{x}) > \xi \Rightarrow$ Reject H_0



Neyman-Pearson Lemma: Given $L(X), \xi$ so that $P[L(X) > \xi \mid H_0] = \alpha$ and $P[L(X) \leq \xi \mid H_1] = \beta$, then for any other test (rejection region) w/ $P[X \in R \mid H_0] \leq \alpha$, then $P[X \notin R \mid H_1] \geq \beta$.
Sig. Testing: Given $X_1, \ldots, X_n$, find a rejection reg. so a level of T1 err. is achieved: $P[\text{Reject } H_0 \mid H_0] = \alpha$.
*α: Significance level, $1 - \alpha$: Confidence level.
Bayesian Hyp. Testing: **MAP Rule:** Selects hyp. w/ higher a posteriori prob, reject H_0 if:
$p(H_1 \mid \underline{x}) \underset{H_0}{\overset{H_1}{\gtrless}} p(H_0 \mid \underline{x})$ | $f(H_1 \mid \underline{x}) \underset{H_0}{\overset{H_1}{\gtrless}} f(H_0 \mid \underline{x})$
$p(\underline{x} \mid H_1)\pi_j \underset{H_0}{\overset{H_1}{\gtrless}} p(\underline{x} \mid H_0)\pi_0$ | $f(\underline{x} \mid H_1)\pi_j \underset{H_0}{\overset{H_1}{\gtrless}} f(\underline{x} \mid H_0)\,\pi_0$
*$p(H_j \mid \underline{x}) = \frac{p_{\underline{X}}(\underline{x}|H_j) P[H_j]}{p_{\underline{X}}(\underline{x}|H_0) P[H_0] + p_{\underline{X}}(\underline{x}|H_1) P[H_1]}$: A posteriori
Min. Cost Bayes' Dec. Rule: $C_{i,j}$ is cost of accepting H_j when H_i is in place, so the MCBDR minimizes the avg. cost
Min. Cost Detection $= \sum_{i=0}^{1} \sum_{j=0}^{1} C_{ij} \; P[\text{decide } j \mid H_i] \; \pi_i$
*$j = 0$: Accept H_0, $j = 1$: Reject H_0
Min. Cost Dec. Rule: Given $\Lambda(\underline{x}) = \frac{f_{\underline{X}}(\underline{x}|H_1)}{f_{\underline{X}}(\underline{x}|H_0)}$, then
Accept H_0 if $\Lambda(\underline{x}) < \frac{\pi_0(C_{01} - C_{00})}{\pi_1(C_{10} - C_{11})}$.
Accept H_1 if $\Lambda(\underline{x}) \geq \frac{\pi_0(C_{01} - C_{00})}{\pi_1(C_{10} - C_{11})}$.
Naive Bayes Assumption: Assume $X_1 \ldots, X_n$ (features) are ind., then $p_{\underline{X}|\Theta}(\underline{x} \mid \theta) \Pi_{i=1}^{n} p_X(x_i \mid \theta)$.
Notation: $P_{\underline{X}|\Theta}(\underline{x}|\theta)$, only put RVs in subscript, not values.
$P_{\underline{X}}(\underline{x}|H_i)$, didn't put H in subscript b/c it's not a RV.