Probability

Bernoulli Random Variable

Grading: 30% Assignments, 25% Midterm (Oct 22th), 45% Final (Dec 12th - 19th)

Geometric Random Variable

* Memoryless:

Exponential Random Variable

* Memoryless:

Probability Fundamentals

1. **Models and Basic Theorems**:
   The basic setup for probability is intuitive, main concepts involves:
   sample space, counting, probability laws and axioms, additivity

2. **Conditioning**:
   Definition: $P(A|B) = \frac{P(A \cap B)}{P(B)}$

   Multiplication rule: $P(A \cap B) = P(B)P(A|B)$
   Total probability theorem: $P(B) = \sum_i P(A_i) \cdot P(B|A_i)$
   Bayes' Rule: $P(A_i|B) = \frac{P(A_i)P(B|A_i)}{\sum_i P(A_i)P(B|A_i)}$

3. **Independence**:
   Definition: $P(A \cap B) = P(A) \cdot P(B)$
   The occurrence (or not) of an event give no extra info.

   Conditional independent: $P(A \cap B|C) = P(A|C) \cdot P(B|C)$
   Collection of event: $P(A_i \cap A_j ... \cap A_m) = \Pi_{k=i}^{m} P(A_k)$
   Pairwise independent: $\not\rightarrow$ independent

4. **Combination and Permutation**:
   Permutation: $n!$
   Combination: $\binom{n}{k} = \frac{n!}{k!(n-k)!}$
   Sum of combination: $\sum \binom{n}{k} = 2^n$

Random Variables

1. **Discrete Random Variables**:
   PMF: Probability Mass Function
   Bernoulli, Uniform, Binomial, Geometric, Poisson
   Expectation & Variance
   Conditional PMF
   Joint and Margin PMF for multiple variables

2. **Continuous Random Variables**
   PDF: Probability Density Function
   CDF: Cumulative Density Function
   Uniform, Exponential, Normal
   Conditional PDF: $f_{X|X \in A}(x) = \frac{f_X(x)}{P(A)}$ if $x \in A$
   Joint and Margin PDF for multiple variables:
   $F_{X,Y}(x,y) = \int_{-\infty}^{y} [\int_{-\infty}^{x} f_{x,y}(s,t)ds]\ dt$
   $f_{X,Y}(x,y) = \frac{\partial^2 F_{X,Y}}{\partial x \partial y}(x,y)$

3. **Derived Distribution**
   Discrete: $Y = g(X), P(Y = y) = P(X = g^{-1}(Y))$
   Ex: $Y = aX + b, P(Y = y) = P(X = \frac{y-b}{a})$
   Continuous: $f_Y(y) = |\frac{g^{-1}(y)}{dy}| f_X(g^{-1}(y))$
   Ex: $Y = aX + b, f_Y(y) = \frac{1}{|a|} f_X(\frac{y-b}{a})$

   General Procedure:
   a). Find CDF
   b). Differentiate to find PDF

4. **Sum of random variables**
   $Z = X + Y$, ——— X, Y are independent
   Discrete: $p(z) = \sum_x p_X(x) \cdot p_Y(z - x)$
   Continuous: $f(z) = \int_{-\infty}^{\infty} f_X(x) \cdot f_Y(z - x)dx$

5. **Covariance & Correlation**
   $cov(X,Y) = E[(X - E[X])(Y - E[Y])]$
   $cov(X,Y) = E[XY] - E[X]E[Y]$
   $var(X_1 + ... + X_n) = \sum_1^n var(X_i) + \sum_{i \neq j} cov(X_i, X_j)$
   $\rho_{X,Y} = \frac{cov(X,Y)}{\sigma_X \sigma_Y}$

6. **Law of total variance**
   $Var[X] = E[Var[X|Y]] + Var[E[X|Y]]$
   (variability within group + variability between groups)

7. **Law of iterative expectation**
   $E[X] = E[E[X|Y]]$

8. **Sum of random number of independent variables**
   X: the variable to be sumed
   N: the number of X being sumed, N is a random variable
   X and N are independent
   Y: $Y = X_1 + X_2 + ... + X_N$
   $E[Y] = E[N]E[X]$
   $Var[Y] = E[N]Var(X) + E[X]^2 Var(N)$

Bayesian Statistics

1. **Basic Setup**:
   a). Prior, Evidence, Posterior
   b). MAP (gives smallest errors), LMS
   c). Performance: conditional / overall prob of error

2. **Recognizing Normal**:
   a). Recognizing Normal PDF $f_X(x) = c \cdot e^{-(\alpha x^2 + \beta x + \gamma)}$
   b). The mean is $-\frac{\beta}{2\alpha}$, the variance is $\frac{1}{2\alpha}$

3. **Linear Model: Normal + Additive Noise**
   a). $X = \Theta + W$, $\Theta, W \sim N(0,1)$, independent
   b). Given observation of X, to estimate $\theta \rightarrow f_{\Theta|X}(\theta|x)$
   c). By to Bayes' Rule :$f_{\Theta|X}(\theta|x) = \frac{f_\Theta(\theta) \cdot f_{X|\Theta}(x|\theta)}{f_X(x)}$
   d). For each term: $f_{X|\Theta}(x|\theta) \rightarrow X = \theta + W \sim N(\theta, 1)$
   f). Applying all: $f_{\Theta|X}(\theta|x) = \frac{c_\theta \cdot e^{-\frac{1}{2}\theta^2} \cdot c_x \cdot e^{-\frac{1}{2}(x-\theta)^2}}{f_X(x)}$

   In summary: $f_{\Theta|X}(\theta|x) = c(x) \cdot e^{\text{quad}(\theta)}$, hence is normal distribution, both MAP and LMS give estimator $\frac{x}{2}$

4. **Linear Model: multiple observations**
   $X_1 = \Theta + W_1$,
   $X_2 = \Theta + W_2$,
   ...
   $X_n = \Theta + W_1$, $\Theta \sim N(x_0, \sigma_0^2), W_i \sim N(0, \sigma_i^2)$

   As we can see, for each $f_{X_i|\Theta}(x_i|\theta)$ is still normal
   Joint distribution: $f_{X_1,X_2...X_n|\Theta}(x_1,x_2...x_n|\theta) \sim N$
   Using base rule, the posterior should be also normal
   It is easy to verify that: estimator $= \frac{\sum_{i=0}^{n} \frac{x_i}{\sigma_i^2}}{\sum_{i=0}^{n} \frac{1}{\sigma_i^2}}$
   Take away: observations, including prior, are weighted

   Performance Measure:
   $E[(\Theta = \hat{\Theta})|X = x] = E[(\Theta = \hat{\theta})|X = x]$
   $= Var[\Theta|X = x] = \frac{1}{\sum_{i=0}^{n} \frac{1}{\sigma_i^2}} = E[(\Theta = \hat{\Theta})]$
   Take away: not related to observation!

5. **Multiple Parameters** $\theta_0, \theta_1, \theta_2 ...$
   a). Get priors of the parameters, and joint distribution
   b). Use Bayes' Rule to write posterior

6. **Linear Normal Model:**
   The posterior is normally distributed

7. **Leat Square Error:**
   1. Without any observation:
   To get LSE: $E[(\Theta - \hat{\theta})^2]$, we have $\hat{\theta} = E[\Theta]$
   MSE: $E[(\Theta - \hat{\theta})^2] = Var[(\Theta - \hat{\theta})] + E[(\Theta - \hat{\theta})]^2 = Var[\Theta]$

   2. With one observation:
   To get LSE: $E[(\Theta - \hat{\theta})^2 | X = x]$, we have $\hat{\theta} = E[\Theta|X = x]$
   MSE: $E[(\Theta - \hat{\theta})^2 | X = x] = Var[\Theta|X = x]$

   3. LSE With multiple observations or unknowns
   $\hat{\theta}_j = E[\theta_j | X_1 = x_1, ..., X_n = x_n]$

   Take Away: Property of LSE:
   Let $\tilde{\Theta} = \hat{\Theta} - \Theta$ be the error of the LSE:
   $E[\tilde{\Theta}] = 0$; $Cov[\tilde{\Theta}, \hat{\Theta}] = 0$; $Var[\Theta] = Var[\hat{\Theta}] + Var[\tilde{\Theta}]$

8. **Linear Leat Square Error:**
   Since LSE sometimes is hard to compute, we may want to constraint our estimator form to be linear: $\hat{\Theta} = aX + b$

   We have the $\hat{\Theta}_L$ to be: min $E[(\Theta - aX - b)^2]$, wrt a,b
   Set $b = E[\Theta] - aE[X]$, then solve a to be: $\frac{Cov(\Theta, X)}{Var(X)}$
   $E[\Theta] + \frac{\text{Cov}(\Theta,X)}{\text{Var}(X)}(X - E[X]) = E[\Theta] + \rho \frac{\sigma_\Theta}{\sigma_X}(X - E[X])$
   MSE: $E[(\hat{\Theta}_L - \theta)^2] = (1 - \rho^2) Var[\Theta]$

   Multiple Observation: min: $E[(\Theta - a_1 X_1 - ... a_n X_n - b)^2]$
   Apply the same logic of how to derive b and a, solve the linear equation and get the result.

King’s Sibling

The king comes from a family of 2 children, what is the probability that his sibling is female?

Expected trails to flip 3 heads in a row

Test