Random Variables

Random Variable: A variable with a stochastic outcome

\[ X = x \quad x \in \{1, 2, 3, 4, 5, 6\} \]

Event: A set of outcomes

\[ X \geq 3 \rightarrow \{3, 4, 5, 6\} \]

Probability: The chance that an event occurs

\[ P(X \geq 3) = \frac{4}{6} \]
Distributions

A distribution maps outcomes to probabilities

\[ P(X=x) = \frac{1}{6} \]

Commonly used (or abused) shorthand:

\[ P(x) \leftrightarrow P(X=x) \]
Conditional Probabilities

Joint Probability

\[ P(A, B) := P(A \cap B) \]

\[ \uparrow \quad \uparrow \]

Events

Conditional Probability

\[ P(A \mid B) := \frac{P(A, B)}{P(B)} \]
Sum Rule

General Case

\[ 0.5 + 0.6 - 0.4 = 0.7 \]

\[ P(A \cup B) = P(A) + P(B) - P(A,B) \]

Either a man or has short hair

0.5

0.6

Marginal

Corrollaries

\[ P(A) = \sum_{x \in A} P(X=x) \]

\[ P(Y=y) = \sum_{x} P(Y=y, X=x) \]
Bayes' Rule

\[ P(A \mid B) = \frac{P(A, B)}{P(B)} \]

Product Rule
\[
P(A, B) = P(A \mid B) P(B)
= P(B \mid A) P(A)
\]

Bayes' Rule
\[
P(A \mid B) = \frac{P(A, B)}{P(B)}
= \frac{P(B \mid A) P(A)}{P(B)}
\]
Example

Event

Prior

A: You have a rare disease

\[ P(A) = 0.0001 \quad P(\neg A) = 0.9999 \]

B: Test for disease is positive

\[ P(B \mid A) = 0.99 \quad P(B \mid \neg A) = 0.01 \]

Question: What is \( P(A \mid B) \)?

\[
P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B)} = \frac{0.0001}{0.01 + 0.0001} \approx 0.01
\]
Probability Densities

Suppose that $X$ is a continuous variable then $P(X=x)$ is 0 for any outcome $x$.

$X \sim \text{Normal}(10,1)$

$p(X=10) = 0$

$p(3 \leq X \leq 4) \neq 0$

Define density function

$$p_X(x) = \lim_{\delta \to 0} \frac{P(x-\delta \leq X < x+\delta)}{2 \delta}$$
Probability Space \((\Omega, F, P)\)

\(\Omega\) Sample space
(set of possible outcomes)

\(F\) Set of events
(every possible subset of the sample space)

\(P\) Probability measure

\[ P : F \to [0, 1] \]

Event Probability

\[ P(\emptyset) = 0 \] (Empty set has prob 0)

\[ P(\Omega) = 1 \] (Sample space has prob 1)

\[ P\left( \bigcup E_i \right) = \sum P(E_i) \]

when \(\{E_i\}\) disjoint
Examples of Reference Measures (not probability)

Lebesgue Measure:
\[ \mu([a, b]) = b - a \] (Width of interval)

Counting Measure:
\[ \mu(\{x_i \in \mathbb{N} \}) = \# \] (Number of elements)

Product Measure:
\[ \mu(E) = \mu_1(E_1) \mu_2(E_2) \] (Cartesian Product)
\[ E = (E_1, E_2) \]
Definition of Probability Measure

Probability measure

\[ P(A) := \int_{x \in A} P_X(x) \, d\mu_X(x) \]

Event Outcome

Differential of the net measure

Density function

Machine Learning Notation

\[ P(A) = \int_A P(x) \, dx \]

Ref measure implied
Expected Values

\[ E[X] := \int x \, p(x) \, dx \]

\[ X \sim p(x) \]

Implied by (1) Statistician defines this first

Conditional Expectation

\[ E[f(x, y) \mid y = y] := \int f(x, y) \, p(x \mid y) \, dx \]

Observed data

Expectation w.r.t. a different distribution

\[ E_{q(x)} [f(x)] = \int f(x) \, q(x) \, dx \]

↑ some dist q(x)
Central problem in this course

\[ E_p(x,y) [f(x,y)] \]

↑ Things we do know
↓ Things we don't know

Examples

Self-driving Cars

\( y \) Past trajectory
\( x \) Future trajectory
\( f \) Will pedestrian cross?

Diagnosis

Symptoms
Condition
Treatment Outcome
Example: Biased Coins

\[ X \sim \text{Beta}(\alpha, \beta) \]

\[ Y_n \sim \text{Bernoulli}(X) \quad n = 1, \ldots, N \]

Bayes Rule

\[
p(X \mid Y_1 = y_1, \ldots, Y_N = y_N) = \frac{p(Y_1 = y_1, \ldots, Y_N = y_N \mid X) p(X)}{p(Y_1 = y_1, \ldots, Y_N = y_N)} \]

Posterior

Prior

Likelihood

Marginal Likelihood
Prior

\[ \text{Beta}(x; \alpha, \beta) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1} (1-x)^{\beta-1} \]

\[ B(\alpha, \beta) = \frac{\Gamma(\alpha) \Gamma(\beta)}{\Gamma(\alpha+\beta)} \]

Likelihood

\[ p(y_1, \ldots, y_n | x) = \prod_{n=1}^{N} p(y_n | x) \]

\[ p(y_n | x) = \begin{cases} x & y_n = 1 \\ (1-x) & y_n = 0 \end{cases} \]

(Trials are i.i.d.)

\[ x^{y_n} (1-x)^{(1-y_n)} \]
Conjugacy

\[
p(x \mid y_{1:n}) = \frac{p(y_{1:n} \mid x) \cdot p(x)}{p(y_{1:n})} \propto p(y_{1:n} \mid x)
\]

\[
p(y_{1:n}, x) = p(x) \cdot p(y_{1:n} \mid x)
\]

\[
\frac{1}{B(\alpha, \beta)} \cdot \prod_{n=1}^{N} \frac{\alpha - 1}{(1 - x)} \cdot \frac{\beta - 1}{(1 - y_{n})} \\
\cdot \prod_{n=1}^{N} \frac{(1 - y_{n})}{(1 - x)} \\
\cdot \prod_{n=1}^{N} \frac{y_{n}}{(1 - y_{n}) + \alpha - 1} \\
\cdot \prod_{n=1}^{N} \frac{(1 - y_{n})}{(1 - x) + \beta - 1}
\]

Sufficient statistics

- Number of heads in \( N \) trials
- Number of tails in \( N \) trials
Conjugacy

\[ p(x \mid y_1:n) = \frac{p(y_1:n, x)}{p(y_1:n)} \]

\[ p(y_1:n, x) = p(x) p(y_1:n \mid x) \]

\[ \alpha = \frac{\sum_{n=1}^{N} y_n + \alpha}{B(\alpha, \beta)} \]

\[ \bar{\beta} = \frac{\sum_{n=1}^{N} (1-y_n) + \beta}{B(\alpha, \beta)} \]

\[ \alpha \sim \text{Beta}(\alpha + \sum_{n=1}^{N} y_n, \beta) \]

\[ \bar{\beta} \sim \text{Beta}(\alpha, \beta + \sum_{n=1}^{N} (1-y_n)) \]

\[ \mathbb{E}[\alpha] = \frac{\alpha}{\beta} \]

\[ \mathbb{E}[\bar{\beta}] = \frac{\alpha}{\beta} \]

\[ \text{Var}[\alpha] = \frac{\alpha}{\beta^2} \]

\[ \text{Var}[\bar{\beta}] = \frac{\alpha}{\beta^2} \]

\[ \text{Cov}[\alpha, \bar{\beta}] = 0 \]
Conjugacy

\[ p(y_1:n, x) = \frac{\text{Beta}(\tilde{\alpha}, \tilde{\beta})}{\text{Beta}(x; \tilde{\alpha}, \tilde{\beta})} \]

\[ \text{Posterior} \]

\[ p(x | y_1:n) = \text{Beta}(x; \tilde{\alpha}, \tilde{\beta}) \]

\[ \text{Marginal Likelihood} \]

\[ p(y_1:n) = \frac{\text{Beta}(\tilde{\alpha}, \tilde{\beta})}{\text{Beta}(\alpha, \beta)} \]

\[ \tilde{\alpha} = \sum_{n=1}^{n} y_n + \alpha \]

\[ \tilde{\beta} = \sum_{n=1}^{n} \left[ 1 - y_n \right] + \beta \]
Predictive Distribution

Joint probability of next trial and coin bias

\[ p(y_{n+1} \mid y_{1:n}) = \int dx \ p(y_{n+1}, x \mid y_{1:n}) \]

\[ = \int dx \ p(y_{n+1} \mid x) p(x \mid y_{1:n}) \]

\[ = E_{p(x \mid y_{1:n})} \left[ p(y_{n+1} \mid x) \right] \]

Weighted Coin Example

(Exercise)
Why is Bayesian Inference Hard?

Example: Mixture of Gaussians

Center $\mu_h \sim \text{Normal}(0, 1)$ \hspace{1cm} $h = 1, \ldots, K$

Width $\sigma_h \sim \text{Gamma}(1, 1)$

Cluster Assignment $Z_n \sim \text{Discrete}(1/K, \ldots, 1/K)$ \hspace{1cm} $n = 1, \ldots, N$

Assignment $y_n | Z_n = h \sim \text{Normal}(\mu_h, \sigma_h)$

Marginal Likelihood

$p(y_1:n) = \int d\mu_1:k d\sigma_1:k dZ_1:n p(y_1:n, Z_1:n, \mu_1:k, \sigma_1:k)$