Basic Probability Recitation

Dan Schwartz (Instructor), HMW-Alexander (Noter)

January 20, 2017

Back to Index

Contents

Resources

- Lecture

---

1 Events, Evnet Spaces

- Events (ω) are possible outcomes of a random experiment.
- An event space (Ω) is the set of all possible outcomes.
- $\Omega = \{\omega_0, \omega_1, \ldots, \omega_n\}$

2 Random variables

Random variables are functions from events to real numbers:

$$X : \Omega \to \mathbb{R}$$

3 Probability

- Probability measure is in reference to a subset of event outcomes occurring.
- A function from subsets to $[0, 1]$
- Notation:
  - $P(X)$ often means $P(X = x)$
  - $P(A)$ can mean $P(\omega \in A)$, where $A \subseteq \Omega$
- Axioms
  - $P(\Omega) = 1$
  - $P(\emptyset) = 0$
  - If A disjoint from B, $P(A \cup B) = P(A) + P(B)$

4 Distributions

- Discrete: probability mass function (pmf) gives $P(\{\omega_i\})$ for each outcome.
- Continuous:
  - Cumulative distribution function (cdf), denoted $F(t)$, gives $P(X \leq t)$ for $t \in \mathbb{R}$, $F(-\infty) = 0, F(\infty) = 1$.
  - If there exists f, such that $\int_{-\infty}^{t} f(x)dx = P(X \leq t)$, then f is called the probability density function (pdf).
- Cumulative distribution function applies to both discrete and continuous event spaces.

4.1 Median, Mode

- Median: $t : F(t) = 0.5, P(X \leq t) = P(X > t)$
- Mode: point where pmf or pdf is maximum.

4.2 Mean, Variance

- Mean (expected value): weighted average of $X(\omega)$, where the weights are given by the probability measure.
  - pmf: $E[X] = \sum_i x_i P(X = x_i)$
  - pdf: $E[g(X)] = \int g(x)f(x)dx$
- Variance: $E[(X - E[X])^2]$, how far do values tend to be from the mean, measure of dispersion.

4.3 Joint Distribution

Multidimensional event space, consider an event to be an outcome for all of the variables jointly.

$$P_{XY}(X, Y)$$

4.4 Marginal Distribution

- pmf: $P_X(X) = \sum_y P_{XY}(X, Y = y)$
- pdf: $f_X = \int_y f_{XY}(x, y)dy$

4.5 Independence

X, Y are independent iff $\forall x, y, P_{XY}(X, Y) = P_X(X)P_Y(Y)$

4.6 Mean, Covariance

- Mean: if $Z = (X, Y)$, then $\mathbb{E}[Z] = (\mathbb{E}[X], \mathbb{E}[Y])$ and $\mathbb{E}[g(Z)] = \int_x \int_y g((x,y)) f_{XY}(x,y)dxdy$
- Covariance: $\mathbb{E}[(Z - \mathbb{E}[Z])(Z - \mathbb{E}[Z])^T]$

4.7 Conditional Distributions, Bayes Rule

$$P(X|Y) = \frac{P(X,Y)}{P(Y)} = \frac{P(Y|X)P(X)}{P(Y)} = \frac{P(Y|X)P(X)}{\sum_x P(Y|X)P(X)}$$

4.8 Prior, Likelihood, Posterior

$$P(\Theta|D) = \frac{P(D|\Theta)P(\Theta)}{P(D)}$$

- Prior: $P(\Theta)$, the probability of parameters θ.
- Likelihood: $P(D|\Theta)$, the conditional probability of observing a feature value of d given that the parameters θ.
- Posterior: $P(\Theta|D)$, the conditional probability of correct parameters being θ, given that feature value d has been observed.

5 Distribution Families

- Bernoulli distribution (binary distribution)
  - $Bern(x|\mu) = \mu^x(1-\mu)^{1-x}$, where $x \in \{0, 1\}, 0 \leq \mu \leq 1$
  - $\mathbb{E}[x] = \mu$
  - $var[x] = \mu(1-\mu)$
- Beta distribution
  - $Beta(\mu|a,b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\mu^{a-1}(1-\mu)^{b-1}$, where $\Gamma(x) = \int_0^\infty u^{x-1}e^{-u}du$
  - $\mathbb{E}[x] = \frac{a}{a+b}$
  - $var[x] = \frac{ab}{(a+b)^2(a+b+1)}$
- Multinomial distribution
  - $p(\vec{x}|\vec{\mu}) = \prod_{k=1}^K \mu_k^{x_k}$, where $\vec{x}$ is encoded with $\{0, 1\}$, and $\sum_k x_k = 1$, $\sum_k \mu_k = 1$
  - $\mathbb{E}[\vec{x}|\vec{\mu}] = \vec{\mu}$
- Dirichlet distribution
  - $Dir(\vec{\mu}|\vec{\alpha}) = \frac{\Gamma(\alpha_0)}{\Gamma(\alpha_1)...\Gamma(\alpha_K)} \prod_{k=1}^K \mu_k^{\alpha_k - 1}$, where $\alpha_0 = \sum_{k=1}^K \alpha_k$
- Gaussian distribution
  - $\mathcal{N}(\vec{x}|\vec{\mu}, \Sigma) = \frac{1}{(2\pi)^{D/2}|\Sigma|^{1/2}} \exp\{-\frac{1}{2}(\vec{x} - \vec{\mu})^T \Sigma^{-1} (\vec{x} - \vec{\mu})\}$