MATH 728D: Machine Learning Solutions to Homework #2: Probability

John Burkardt

February 17, 2019

1 Properties of Probability Measures

Assume that P is a probability measure defined on a sample space Ω, that A and B are events (measurable subsets of Ω), and that A^C represents the event that A *does not* occur. Verify the following statements:

1. $P(A^C) = 1 - P(A)$;

$$\begin{aligned} \Omega &= A + A^C \\ P(\Omega) &= P(A) + P(A^C) \\ 1 &= P(A) + P(A^C) \\ P(A^C) &= 1 - P(A). \end{aligned}$$

2. $P(B \cap A^C) = P(B) - P(A \cap B)$;

$$\begin{aligned} B &= (B \cap A) + (B \cap A^C) \\ P(B) &= P(B \cap A) + P(B \cap A^C) \\ P(B \cap A^C) &= P(B) - P(B \cap A). \end{aligned}$$

3. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$;

$$\begin{aligned} A \cup B &= (A \cap B) + (A \cap B^C) + (A^C \cap B) \\ P(A \cup B) &= P(A \cap B) + P(A \cap B^C) + P(A^C \cap B) \\ P(A \cup B) &= P(A \cap B) + P(A \cap B^C) + P(A^C \cap B) + P(A \cap B) - P(A \cap B) \\ P(A \cup B) &= P(A) + P(B) - P(A \cap B) \end{aligned}$$

4. $P(\cup_{i \in I} A_i) \leq \sum_{i \in I} P(A_i)$;
   Define $B_1 = A_1, B_2 = A_2 \cap A_1^C, B_3 = A_3 \cap A_1^C \cap A_2^C$, and so on;
   Then $P(B_i) \leq P(A_i)$ because $B_i \subseteq A_i$ for every i. Also, the B_i sets are disjoint.

$$\begin{aligned} \cup_{i \in I} A_i &= \cup_{i \in I} B_i \\ P(\cup_{i \in I} A_i) &= P(\cup_{i \in I} B_i) \\ P(\cup_{i \in I} B_i) &= \sum_{i \in I} P(B_i) \leq \sum_{i \in I} P(A_i) \\ P(\cup_{i \in I} A_i) &\leq \sum_{i \in I} P(A_i) \end{aligned}$$

5. $P(A) = \sum_{i \in I} P(A \cap C_i)$, if $\{C_i\}$ is a partition of Ω;
   The set A is the disjoint sum of the sets $(A \cap C_i)$. Therefore,

$$\begin{aligned} A =& \cup_{i \in I} (A \cap C_i) \\ P(A) =& P(\cup_{i \in I} (A \cap C_i)) \\ P(A) =& \cup_{i \in I} P(A \cap C_i)) \text{ (disjoint sum)} \end{aligned}$$

6. $A \subseteq B \Rightarrow P(A) \leq P(B)$;

$$\begin{aligned} B =& A + (B \cap A^C) \\ P(B) =& P(A) + P(B \cap A^C) \\ P(B \cap A^C) \geq& 0 \\ P(B) \geq& P(A) \end{aligned}$$

2 Roulette

1. What is the probability that S is red?
   There are 38 slots, of which 18 are red, and all slots are equally likely to be chosen. Therefore, the chance that the ball will land in a red slot is $P(\text{red}) = \frac{18}{38} \approx 0.4737$.
2. What is the probability that S lands in a slot numbered between 1 and 10?
   There are 38 slots, of which 10 are numbered between 1 and 10, and all slots are equally likely to be chosen. Therefore, the chance of that the ball will land in such a slot is $P(\text{1-to-10}) = \frac{10}{38} \approx 0.2632$.
3. What is the probability that S lands on slot #7?
   $P(7) = \frac{1}{38} = 0.0263$;
4. If the player has bet B=$100 on #7, what is the expected value of the game?
   $EV = P(7) * \$35 * 100 + P(\text{not } 7) * -\$100 = -\$5.26$

3 Chuck-a-Luck

1. What is the probability of exactly 0, 1, 2 or 3 D's showing up? (Note that 1 or 2 D's can show up in 3 different ways!)

$$\begin{aligned} P(0) =& 1 * \frac{5}{6} \times \frac{5}{6} \times \frac{5}{6} = \frac{125}{216} \approx 0.5787 \\ P(1) =& 3 * \frac{1}{6} \times \frac{5}{6} \times \frac{5}{6} = \frac{75}{216} \approx 0.3472 \\ P(2) =& 3 * \frac{1}{6} \times \frac{1}{6} \times \frac{5}{6} = \frac{15}{216} \approx 0.0694 \\ P(3) =& 1 * \frac{1}{6} \times \frac{1}{6} \times \frac{1}{6} = \frac{1}{216} \approx 0.0046 \end{aligned}$$

2. If the player bets B=$100, what is the expected value of the game?;

$$EV = \$100 * (-1 * \frac{125}{216} + 1 * \frac{75}{216} + 2 * \frac{15}{216} + 10 * \frac{1}{216}) \approx -4.6296$$

3. Suppose we can increase the payoff for the triple D result from $10*B to $N*B. What is the smallest (integer) value of N so that the expected value of the game is in the player's favor (making the expected

value positive)?;
Try $N = 20$:

$$EV = \$100 * (-1 * \frac{125}{216} + 1 * \frac{75}{216} + 2 * \frac{15}{216} + 20 * \frac{1}{216}) \approx 4.1667$$

4 The Game of Life

1. What is the probability that you will live to *at least* age 50?
   Out of 100,000 people who were born, 92,632 are still alive at age 50; this suggests your probability of living to at least 50 is $\frac{92632}{100000} \approx 92.6\%$
2. What is the probability that you will die sometime between age 50 and age 60?
   92,632 people were alive at age 50, and only 85,802 were alive at age 60. If we assume you are alive at age 50, then your chances of dying before 60 can be estimated at $\frac{92632-85802}{92632} = \frac{6830}{92632} \approx 7.4\%$
3. What is the expected lifetime of a person in this population? We want to estimate the area under a graph whose x axis is age, and whose y axis is the probability of reaching that age. The y value is found simply by dividing the number of people by 100,000. Simpson's rule for estimating the area under the function $y(x)$ given n data values at points with equal spacing h is:

$$\text{Area} \approx (\frac{1}{2}y_1 + y_2 + ... + y_{n-1} + \frac{1}{2}y_n) * h$$

Our calculation involves $n = 13$ data values y separated by a uniform spacing of $h = 10$ years:

$$\text{Estimate for Expected Value of Lifespan} =$$
$$(1/2 * 1.00000 + 0.99184 + 0.98771 + 0.97393 + 0.95603 + 0.92632 + 0.85802 + 0.73100$$
$$+ 0.50564 + 0.17915 + 0.00932 + 0.00002 + 1/2 * 0.00000) * 10 \approx 71.13 \text{ years}$$

5 Covariance Definition

(From "Probability Basics" slides):
Let $X = (X_1, ..., X_n)^T, Y = (Y_1, ..., Y_n)^T$ be vectors of random variables with joint distribution P. The covariance of X and Y is the rank-one matrix

$$\text{cov}[X, Y] := \mathbb{E}[(X - \mathbb{E}[X]) (Y - \mathbb{E}(Y))^T] = \mathbb{E}[X\, Y^T] - \mathbb{E}[X]\, \mathbb{E}[Y]^T$$

Verify the second equality.

$$\begin{aligned} \text{cov}[X, Y] : &= \mathbb{E}[\, (X - \mathbb{E}[X]) \quad (Y - \mathbb{E}(Y))^T\,] \\ &= \mathbb{E}[X\, Y^T - X\, \mathbb{E}[Y]^T - \mathbb{E}[X]\, Y^T + \mathbb{E}[X]\, \mathbb{E}[Y]^T] \\ &= \mathbb{E}[X\, Y^T] - \mathbb{E}[X\, \mathbb{E}[Y]^T] - \mathbb{E}[\mathbb{E}[X]\, Y^T] + \mathbb{E}[\mathbb{E}[X]\, \mathbb{E}[Y]^T] \\ &= \mathbb{E}[X\, Y^T] - \mathbb{E}[X]\, \mathbb{E}[Y]^T - \mathbb{E}[X]\, \mathbb{E}[Y^T] + \mathbb{E}[X]\, \mathbb{E}[Y]^T \\ &= \mathbb{E}[X\, Y^T] - \mathbb{E}[X]\, \mathbb{E}[Y]^T \end{aligned}$$

6 Tail Bounds

(From "Probability Basics" slides):

1. Show that, for any nonnegative random variable X:

$$\mathbb{E}[X] = \int_0^\infty \text{Prob}(X \geq t)\, dt$$

$$\begin{aligned}
\text{Prob}(X \geq t) &= \int_{X=t}^\infty p(X)\, dX \\
\int_{t=0}^\infty \text{Prob}(X \geq t)\, dt &= \int_{t=0}^\infty \int_{X=t}^\infty p(X)\, dX\, dt \\
&= \int_{X=0}^\infty \int_{t=0}^X p(X)\, dt\, dX \\
&= \int_{X=0}^\infty X\, p(X)\, dX \\
&= \mathbb{E}(X)
\end{aligned}$$

and re-derive Markov's inequality:

$$\begin{aligned}
\mathbb{E}[X] &= \int_{t=0}^a \text{Prob}(X \geq t)\, dt + \int_{t=a}^\infty \text{Prob}(X \geq t)\, dt \\
&\geq \int_{t=0}^a \text{Prob}(X \geq t)\, dt \\
&\geq \int_{t=0}^a \text{Prob}(X \geq a)\, dt \quad (\text{Prob}(X \geq t) \text{ is non increasing}) \\
&\geq a\, \text{Prob}(X \geq a)
\end{aligned}$$

2. Let $\phi(t)$ be any strictly monotonically increasing nonnegative function. Show that, for any random variable X and any $t \in \mathbb{R}$:

$$\text{Prob}(X \geq t) \leq \frac{\mathbb{E}[\phi(X)]}{\phi(t)}$$

Since $\phi(t)$ is monotonic, we have

$$\text{Prob}(X \geq t) = \text{Prob}(\phi(X) \geq \phi(t))$$

Now apply Markov's inequality with $Y = \phi(X)$ and $a = \phi(t)$:

$$\text{Prob}(X \geq t) = \text{Prob}(\phi(X) \geq \phi(t)) \geq \frac{\mathbb{E}[\phi(X)]}{\phi(t)}$$

Alternately:

$$
\begin{aligned}
\mathbb{E}[\phi(X)] =& \int_{X=0}^{\infty} \phi(X)\, p(X)\, dX \\
=& \int_{X=0}^{t} \phi(X)\, p(X)\, dX + \int_{X=t}^{\infty} \phi(X)\, p(X)\, dX \\
\geq& \int_{X=t}^{\infty} \phi(X)\, p(X)\, dX \\
\geq& \int_{X=t}^{\infty} \phi(t)\, p(X)\, dX \\
=& \phi(t) \int_{X=t}^{\infty} p(X)\, dX \\
=& \phi(t)\, \text{Prob}(t <= X)
\end{aligned}
$$

from which the conclusion follows.

3. From the previous result,

$$
\text{Prob}(X \geq t) \leq \frac{\mathbb{E}[\phi(X)]}{\phi(t)}
$$

re-derive Chebychev's inequality that, for an arbitrary random variable X and $t > 0$, one has

$$
\text{Prob}(|X - \mathbb{E}[X]| \geq t) \leq \frac{\text{var}[X]}{t^2}
$$

Note that

$$
\begin{aligned}
\text{var}(X) =& \mathbb{E}[(X - \mathbb{E}(X))^2] \\
=& \text{var}(X - \mathbb{E}(X)) \\
=& \text{var}(|X - \mathbb{E}(X)|)
\end{aligned}
$$

Define $Z(X) = |X - \mathbb{E}[X]|$, and let $\phi(Z) = Z^2$. Then $\phi(Z)$ is a strictly monotonically increasing nonnegative function, and we write

$$
\begin{aligned}
\text{Prob}(|X - \mathbb{E}[X] \geq t) =& \text{Prob}(Z \geq t) \\
\leq& \frac{\mathbb{E}[\phi(Z)]}{\phi(t)} \\
=& \frac{\mathbb{E}[(|X - \mathbb{E}[X]|)^2]}{t^2} \\
=& \frac{\text{var}(X)}{t^2}
\end{aligned}
$$

7 Mean and Variance By Sampling

(From "Probability Basics" slides):

Suppose we make N draws from a Gaussian distribution whose true mean is μ and true variance is σ^2.

The maximum likelihood estimates for mean and variance, μ_{ML} and σ^2_{ML}, depend on the random draws X, and are therefore random variables. We can compute the expectation of these quantities. Show that

1. $$\mathbb{E}[\mu_{ML}] = \mu$$

$$\begin{aligned}
\mathbb{E}[\mu_{ML}] =& \mathbb{E}[\frac{1}{N}\sum_{i=1}^{N} x_i] \\
=& \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}[x_i] \\
=& \frac{1}{N}\sum_{i=1}^{N}\mu \\
=& \mu
\end{aligned}$$

2. $$\mathbb{E}[\sigma^2_{ML}] = \frac{(N-1)}{N}\sigma^2$$

The following identities will be used in the argument:

$$\begin{aligned}
\sum_{i=1}^{n}(x_i - \overline{x})^2 =& \sum_{i=1}^{n} x_i^2 - n * \overline{x} \\
\mathrm{var}(t) =& \mathbb{E}[t^2] - (\mathbb{E}[t])^2
\end{aligned}$$

Write:

$$\begin{aligned}
\sigma^2_{ML} :=& \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \mu_{ML})^2 \\
=& \frac{1}{n-1}(\sum_{i=1}^{n} x_i^2 - n * (\overline{x})^2) \\
=& \frac{1}{n-1}\sum_{i=1}^{n} x_i^2 - \frac{1}{n(n-1)} * (\sum_{i=1}^{n} x_i)^2
\end{aligned}$$

Taking the expected value:

$$
\begin{aligned}
\mathbb{E}[\sigma_{ML}^2] =& \mathbb{E}[\frac{1}{n-1}\sum_{i=1}^{n} x_i^2 - \frac{1}{n(n-1)} * (\sum_{i=1}^{n} x_i)^2] \\
=& \frac{1}{n-1}\sum_{i=1}^{n} \mathbb{E}[x_i^2] - \frac{1}{n(n-1)} * \mathbb{E}[(\sum_{i=1}^{n} x_i)^2] \\
=& \frac{1}{n-1}\sum_{i=1}^{n}[\mathrm{var}(x_i) + (\mathbb{E}[x_i])^2] - \frac{1}{n(n-1)} * (\mathrm{var}(\sum_{i=1}^{n} x_i) + (\mathbb{E}[\sum_{i=1}^{n} x_i])^2] \\
=& \frac{1}{n-1}\sum_{i=1}^{n}(\sigma^2 + \mu^2) - \frac{1}{n(n-1)} * (\mathrm{var}(\sum_{i=1}^{n} x_i) + (\sum_{i=1}^{n} \mathbb{E}[x_i])^2) \\
=& \frac{n}{n-1}(\sigma^2 + \mu^2) - \frac{1}{n(n-1)} * ((\sum_{i=1}^{n} \mathrm{var}(x_i)) + (n * \mu)^2) \\
=& \frac{n}{n-1}(\sigma^2 + \mu^2) - \frac{1}{n(n-1)} * (n * \sigma^2 + (n * \mu)^2) \\
=& \frac{n}{n-1}(\sigma^2 + \mu^2) - \frac{1}{n-1} * (\sigma^2 + n * \mu^2) \\
=& \sigma^2
\end{aligned}
$$