B555: Assignment 1

Pawan Patel

September 16, 2015

1. Let $A, B \in \mathcal{F}$. Notice that $P(A \cup B) = P(A \cup (B \cap A^c)) = P(A) + P(B \cap A^c)$, since A, $B \cap A^c$ are disjoint. Likewise, notice that $P(B) = P((A \cap B) \cup (B \cap A^c)) = P(A \cap B) + P(B \cap A^c) \implies P(B \cap A^c) = P(B) - P(A \cap B)$, since $A \cap B$ and $B \cap A^c$ are disjoint. Thus, we have $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

2. The statement is false. Consider a fair coin being flipped twice independently. Let $A =$ the event that the first coin is Heads. Let $B =$ the event that the second coin is Heads. Then, $P(A|B) + P(A|B^c) = P(A) + P(A) = 1 \neq 1/2 = P(A)$

3. a) For player 1 to win, there needs to be four heads before two tails. So he can either win in four flips they're all Heads or in 5 flips if the last one is heads and there are 3Hs and 1 Ts in the first flour. Thus, the probability Player 1 wins is $(1/2)^4 + 4(1/2)^5$.

   b) Notice that player 1 needs $(n - m)$ heads to win and player 2 needs $n - l$ tails to win. In $n - m + n - l - 1$ flips, the game must end as there must then be at least $n - m$ heads or $n - l$ tails. Moreover, in $n - m + n - l - 1$ flips, only one could have won. That is to say, in that many flips, there cannot be both $n - m$ heads and $n - l$ tails. So, player 1 wins if and only if he gets at least $n - m$ heads in the first $n - m + n - l - 1$ flips. This is computed using the binomial formula and summing:

$$P(\text{Player 1 wins}) = \sum_{i=n-m}^{n-m+n-l-1} \binom{n-m+n-l-1}{i} (\frac{1}{2})^{n-m+n-l-1}$$

4. a) Notice:

$$\mu = E[X] = \int_{\mathbb{R}^{d\times d}} x d\mu_X$$
$$= \int_{\mathbb{R}^d} ... \int_{\mathbb{R}^d} (\sum_{i=1}^{n} a_i x_i) P(X_1 = x_1, ..., X_n = x_n)$$
$$= \int_{\mathbb{R}^d} ... \int_{\mathbb{R}^d} a_1 x_1 P(X_1 = x_1, ..., X_n = x_n) + ... + \int_{\mathbb{R}^d} ... \int_{\mathbb{R}^d} a_n x_n P(X_1 = x_1, ..., X_n = x_n)$$
$$= \int_{\mathbb{R}^d} a_1 x_1 P(X_1 = x_1) + ... + \int_{\mathbb{R}^d} a_n x_n P(X_n = x_n)$$
$$= \sum_{i=1}^{n} \int_{\mathbb{R}^d} a_i x_i d\mu_{X_i} = \sum_{i=1}^{n} a_i \int_{\mathbb{R}^d} x_i d\mu_{X_i} = \sum_{i=1}^{n} a_i E[X_i] = \sum_{i=1}^{n} a_i \mu_i \in \mathbb{R}^d$$

b) Notice:

$$Var(X) = E[(X-\mu)^T (X-\mu)]$$
$$= E[(X^T - \mu^T)(X-\mu)] = E[X^T X - \mu^T \mu] = E[X^T X] - \mu^T \mu$$

Thus, we first compute:

$$E[X^T X] = \int_{\mathbb{R}^d} ... \int_{\mathbb{R}^d} x^t x d\mu_X$$
$$= \int_{\mathbb{R}^d} ... \int_{\mathbb{R}^d} (\sum_{i=1}^{n} a_i x_i)^T (\sum_{i=1}^{n} a_i x_i) d\mu_X$$
$$= \int_{\mathbb{R}^d} ... \int_{\mathbb{R}^d} ((\sum_{i=1}^{n} a_i^2 x_i^2) + (\sum_{1\le i<j\le n} 2 a_i a_j x_i x_j)) d\mu_X$$
$$= \int_{\mathbb{R}^d} ... \int_{\mathbb{R}^d} (\sum_{i=1}^{n} a_i^2 x_i^2) d\mu_X + \int_{\mathbb{R}^d} ... \int_{\mathbb{R}^d} (\sum_{1\le i<j\le n} 2 a_i a_j x_i x_j) d\mu_X$$

Where the second term is a linear combination of expectations of products of cross terms, which split as the expectation of the products, as the X_i are independent. Thus, we get:

$$Var[X] = E[X^T X] - \mu^T \mu$$
$$= \int_{\mathbb{R}^d} ... \int_{\mathbb{R}^d} (\sum_{i=1}^{n} a_i^2 x_i^T x_i) d\mu_X + \int_{\mathbb{R}^d} ... \int_{\mathbb{R}^d} (\sum_{1\le i<j\le n} 2 a_i a_j x_i^T x_j) d\mu_X - (\sum_{i=1}^{n} a_i \mu_i)^T (\sum_{i=1}^{n} a_i \mu_i)$$
$$= \sum_{i=1}^{n} (a_i^2 \int_{\mathbb{R}^d} x_i^T x_i P(X_i = x_i)) + (\sum_{1\le i<j\le n} 2 a_i a_j \mu_i^T \mu_j) - (\sum_{i=1}^{n} a_i^2 \mu_i^T \mu_i) - (\sum_{1\le i<j\le n} 2 a_i a_j \mu_i^T \mu_j)$$

(Where, we collect terms here using the fact that $\mu_i^T \mu_j = \mu_j^T \mu_i$)

$$= \sum_{i=1}^{n} a_i^2 E[X_i^T X_i] - \sum_{i=1}^{n} a_i^2 \mu_i^T \mu_i = \sum_{i=1}^{n} a_i^2 (E[X_i^T X_i] - \mu_i^T \mu_i) = \sum_{i=1}^{n} a_i^2 \Sigma_i \in \mathbb{R}^{d\times d}$$

Now, if $Cov[X_1, X_2] = \Lambda$, then the $E[X_1^T X_2]$ would not split above, giving us an extra term at the end of $2a_1a_2(E[X_1^T X_2] - \mu_1^T\mu_2) = 2a_1a_2\Lambda$, that would be added to the variance computed above.

5. a) Let $X_A = 1$ if coin A is heads and 0 otherwise. Define X_B and X_C similarly. Then, $E[X] = E[X_A + X_B + X_C] = E[X_A] + E[X_B] + E[X_C] = .75 + .50 + .25 = 1.5$.

   b) Let X be the event that you flipped 3 Heads and 2 Tails. And, let A, B, and C be the events that you select those coins, respectively. Then,

$$P(C|X) = \frac{P(X|C)P(C)}{P(X|A)P(A) + P(X|B)P(B) + P(X|C)P(C)}$$
$$= \frac{(1/3)\binom{5}{3}(1/4)^3(3/4)^2}{(1/3)\binom{5}{3}(3/4)^3(1/4)^2 + (1/3)\binom{5}{3}(1/2)^5 + (1/3)\binom{5}{3}(1/4)^3(3/4)^2}$$
$$= \frac{(1/4)^3(3/4)^2}{(3/4)^3(1/4)^2 + (1/2)^5 + (1/4)^3(3/4)^2} = \frac{9}{68}$$

6. The statement is True. Consider:

$$\sum_{z\in\Omega} P_{X|YZ}(X = x|Y = y, Z = z)P_{Z|Y}(Z = z|Y = y)$$
$$= \sum_{z\in\Omega} \frac{P_{XYZ}(X = x, Y = y, Z = z)}{P_{YZ}(Y = y, Z = z)} \frac{P_{YZ}(Y = y, Z = z)}{P_Y(Y = y)}$$
$$= \sum_{z\in\Omega} \frac{P_{XYZ}(X = x, Y = y, Z = z)}{P_Y(Y = y)}$$
$$= \frac{P_{XY}(X = x, Y = y)}{P_Y(Y = y)}$$
$$= P_{X|Y}(X = x|Y = y)$$

7. Let X_i be the random variable which is 1 if you roll two sixes on the ith toss of two dice and 0 otherwise. Notice then that X_i is a Bernoulli random varialbe with $p = 1/36$ and that X_i are indepdent in i. Let $X = \sum_{i=1}^{24} X_i$. Then, X is a binomial random variable with $n = 24$ and $p = 1/36$, and it represents the number of times you roll two sixes from 24 tosses of two dice. A player wins if $X \geq 1$, thus we wish to know the $P(X \geq 1)$. We have,

$$P(X \geq 1) = 1 - P(X = 0) = 1 - \binom{24}{0}(1/36)^0(35/36)^{24} = 1 - (\frac{35}{36})^{24}$$

8. P does definte a probability measure. We check the conditions required to be a probability measure.

   i) $0, 1 \in [0, 1] \implies P(\Omega) = P([0, 1]) = 1$

   ii) Let $A_1, A_2, ... \in \mathcal{F}$ be disjoint and Let $A = \cup_{n=1}^{\infty} A_n$. There are four cases:

   (a) If $0, 1 \in A$, then as the A_n are disjoint, we must have $0 \in A_j$ and $1 \in A_k$ for some $j \neq k$. Then, $\sum_{i=1}^{\infty} P(A_n) = P(A_j) + P(A_k) = 1 = P(A)$, as $P(A_n) = 0$ for $n \neq j, k$.

   (b) If $0 \in A$, but $1 \notin A$, then $0 \in A_j$ for a unique j. So, $\sum_{i=1}^{\infty} P(A_n) = P(A_j) = 1/2 = P(A)$, as $P(A_n) = 0$ for $n \neq j$

   (c) Similarly, If $1 \in A$, but $0 \notin A$, then $1 \in A_j$ for a unique j. So, $\sum_{i=1}^{\infty} P(A_n) = P(A_j) = 1/2 = P(A)$, as $P(A_n) = 0$ for $n \neq j$

   (d) Lastly, if $0, 1 \notin A$, then we must have $0, 1 \notin A_j$ for all j. So, $\sum_{i=1}^{\infty} P(A_n) = P(A_j) + P(A_k) = 1 = P(A)$, as $P(A_n) = 0$ for $n \neq j, k$.

   Thus, we have $P(\cup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n)$.

9. a) Notice:

$$P(X = 1, Y = 1, Z = 1) = P(X = 1)P(Y = 1|X = 1)P(Z = 1|X = 1, Y = 1) = ace$$

   And,

$$P(X = 1, Y = 0, Z = 1) = P(X = 1)P(Y = 0|X = 1)P(Z = 1|X = 1, Y = 0) = a(1-c)d$$

   Thus, $P(X = 1, Z = 1) = \sum_y P(X = 1, Y = y, Z = 1) = ace + a(1 - c)d$. Thus,

$$P(Z = 1|X = 1) = \frac{P(Z = 1, X = 1)}{P(X = 1)} = \frac{ace + a(1 - c)d}{a} = d + ce - cd$$

   b) Now, notice:

$$\begin{aligned}
P(Z = 1) = \sum_{x,y} P(Z = 1|X = x, Y = y)P(Y = y|X = x)P(X = x) \\
= eP(X = 1)P(Y = 1|X = 1) + dP(X = 1)P(Y = 0|X = 1) \\
+eP(X = 0)P(Y = 1|X = 0) + dP(X = 0)P(Y = 0|X = 0) \\
= eac + da(1 - c) + e(1 - a)b + d(1 - a)(1 - b) \\
= eac + da - dac + eb - eab + d - db - ad + abd \\
= d + ac(e - d) - ab(e - d) + b(e - d) \\
= d + (e - d)(ac - ab + b) \\
= d + (e - d)(a(c - b) + b)
\end{aligned}$$

10. a) For dim$= 1$ and $\sigma = 1$, we get sample means of 0.244, 0.040, and -0.027 for sample sizes of 10, 100, and 100. For dim$= 1$ and $\sigma = 10$, we get sample means of -0.58, 0.416, and 0.324 for sample sizes of 10, 100, and 1000. We notice the sample means are close to zero, as they should be due to the law of large numbers. But, they are generally slightly further away from zero for samples drawn from $\sigma = 10$ than for $\sigma = 1$.

    b) A covariance matrix equal to the identity means that the components of the random gaussian vector are independent. That is, that X, Y, and Z are independent. If the matrix is changed so that the $Cov(X, Z) = 1$, then X and Z will be highly dependent on each other.

11. Bonus

    a) We first show that as $n \to \infty$, the volume of a hypercube in $\mathbb{R}^n$ is concentrated in the corners. To do this, it suffices to show the ratio of the volume of an n-sphere to a n-hypercube goes to zero as $n \to \infty$. We use the formula for the volume of a n-sphere, which is written with the Gamma function. We recall that Γ grows as $(n-1)!$ Let V_S be the volume of the n-sphere, and V_C be the volume of the n-hypercube. We then have:

$$\lim_{n\to\infty} \frac{V_S}{V_C} = \lim_{n\to\infty} \frac{\frac{n\pi^{n/2}a^{n-1}}{\Gamma(n/2+1)}}{(2a)^n} = \lim_{n\to\infty} \frac{n\pi^{n/2}a^{n-1}}{n!2^n a^n} \to 0$$

    as the factorial in the denominator easily dominates the exponentials in the numerator.

    Next, we show that not only is this volume concentrated in the corners, the corners themselves become long spikes. To do this, it suffices to show that the ratio of the distrance from the center of the cube to the corner to the perpindicular distance to one of the sides is ∞. Let D_c be the distance to the corner and D_s be the distance to the side. We have:

$$\lim_{n\to\infty} \frac{D_c}{D_s} = \lim_{n\to\infty} \frac{\sqrt{na^2}}{a} = \lim_{n\to\infty} \sqrt{n} \to \infty$$

    b) It suffices to show that the ratio of the volume of the thin shell around the n-sphere to the volume of the n-sphere goes to 1 as $n \to \infty$. We have for ϵ very small, but fixed:

$$\lim_{n\to\infty} \frac{\frac{n\pi^{n/2}a^{n-1}}{\Gamma(n/2+1)} - \frac{n\pi^{n/2}(a-\epsilon)^{n-1}}{\Gamma(n/2+1)}}{\frac{n\pi^{n/2}a^{n-1}}{\Gamma(n/2+1)}} = \lim_{n\to\infty} \frac{a^{n-1} - (a-\epsilon)^{n-1}}{a^{n-1}} = \lim_{n\to\infty} 1 - \frac{(a-\epsilon)^{n-1}}{a^{n-1}} \to 1$$