Preliminary Statistics September 2013

Answers to class exercise

Problem Set 3: Probability Models and Distributions

1. (a) Remember that the variance is defined as Var(X) = E[X-μ_x]^2. Given that in the below case (aX-bY) correspond to X and E[X] = μ_x:

$$Var(aX - bY) = E[aX - bY - E(aX - bY)]^2$$

Take the expectation operator through the first bracket applying the rules E[X+Y] = E[X] + E[Y] and E[a] = a, with a being a constant:

$$= E[aX - bY - aE(X) - bE(Y)]^2$$

Rearrange and factorise, so that:

$$= E[a\big(X - E(X)\big) - b(Y - E(Y))]^2$$

Solve the outer bracket by using (a-b)^2 = a^2 + b^2 – 2ab:

$$= E[a^2\big(X - E(X)\big)^2 + b^2\big(Y - E(Y)\big)^2 - 2ab(X - E(X))(Y - E(Y))]$$

From there the expectation operator can be taken through, which gives:

$$= a^2E\big(X - E(X)\big)^2 + b^2E(Y - E(Y))^2 - 2abE\big[\big(X - E(X)\big)\big(Y - E(Y)\big)\big] = a^2Var(X) + b^2Var(Y) - 2abCov(X, Y)$$

Q.E.D.

(b) if X and Y are independent then Cov(X,Y) = 0. From the definition of the covariance we have:

$$Cov(X, Y) = E\big[\big(X - E(X)\big)\big(Y - E(Y)\big)\big]$$

By solving the brackets and taking the expectation operator through one gets:

$$= E[XY - XE(Y) - E(X)Y + E(X)E(Y)]$$

$$= E(XY) - E(X)E(Y) - E(X)E(Y) + E(X)E(Y)$$

$$= E(XY) - E(X)E(Y)$$

From the definition of two independent variables, we know that E(XY) = E(X) E(Y). Hence, the covariance of two independent variables is zero.

2. It is supposed that Marks [M] follow M~N(50,10^2). By standardising (minus the mean divided by the standard deviation; this can also be understood as rescaling the values of the random variable through a linear transformation) one can derive respective values for the standard normal distribution which can then be looked up at any statistical table for the Z-distribution (link).
   a. (i) The probability is found by standardising and exploiting the symmetry of the standard normal distribution, which defines P(Z < -a) = 1 – P(Z<a):
   $$P(M < 30) = P\left(\frac{M-50}{10} < \frac{30-50}{10}\right) = P(Z < -2) = 1 - P(Z < 2) = 0.0228$$
   The probability of a student getting a mark below 30 is 2.28%.

   (ii) The probability is found by standardising and again exploiting the symmetry of the normal (and standard normal) distribution:
   $$P(30 < M < 50) = P\left(\frac{30-50}{10} < \frac{M-50}{10} < \frac{50-50}{10}\right) = P(-2 < Z < 0)$$
   $$= P(Z < 0) - P(Z - 2) = 0.4772$$
   (iii) Solve in exactly the same way (note: the symmetry around zero of the standard normal distribution demand that 50% of the probability lies at each side of 0): $P(M > 50) = P(Z > 0) = 1 - P(Z < 0) = 0.5$

   b. If you want to find $P(M > m^*) = 0.15$ one has to reverse the order of solving the question applied previously, so that: $P(M > m^*) = 0.15 \leftrightarrow P\left(Z > \frac{m^*-50}{10}\right) = 0.15$. From there one can look for a z_i such that the table entry is 0.85 (1-0.15). This is the case for z_i = 1.04. So one knows that P(Z > 1.04) = 0.15. Hence one can solve: $\frac{m^*-50}{10} = 1.04$ with respect to m^*. Hence: P(M > 60.4) = 0.15.
   To get to the top 15% of the distribution, a student needs to get a mark above or equal to 60.4.

3. (a) Write X~N(1,2^2), given that the expected value of a normally distributed random variable is its mean.
   (i) P(|X| >2) can be thought of as the cumulative probability below -2 and above 2 so that: $P(|X| > 2) = P(X < -2) + P(X > 2)$.
   $$P(X < -2) = P\left(Z < \frac{-2-1}{2}\right) = P(Z < -1.5) = 0.0668$$
   $$P(X > 2) = P\left(Z > \frac{2-1}{2}\right) = P(Z > 0.5) = 0.3085$$
   Hence: P(|X| > 2) = 0.0668 + 0.3085 = 0.3753

(ii) P(X < 3|X ≥ 0) can be written as:

$$P(X < 3|X \geq 0) = \frac{P(X < 3) AND P(X \geq 0)}{P(X \geq 0)} = \frac{P(0 \leq X < 3)}{P(X \geq 0)}$$

This is due to the rules for conditional probability. The above equation can then be solved by standardising to find the respective probabilities.

$$\frac{P(0 \leq X < 3)}{P(X \geq 0)} = \frac{P(-0.5 \leq Z < 1)}{P(Z \geq -0.5)} = \frac{0.5328}{0.6915} = 0.77$$

(b) Write $Y \sim t_{20}$ whereby 20 is the degree of freedom (df.). Use a two tailed t-distribution table (link).

The two numbers can be found by exploiting the two tail properties and defining b = -a. The 10% probabilities on each side of the distribution need to be looked up. The critical value for df. 20 and 10% probability is 1.325. Hence P(-1.325 < Y < 1.325) = 0.80 as we know that Y takes with 0.1 probability on values above the critical value of 1.325 and below the critical value of -1.325 each. Hence, realisations of Y lie with 80% probability between this interval [-1,325; 1,325].

(c) Write $W \sim F_{3,10}$ whereby 3 is the df. in the nominator and 10 the df. in the denominator. Us the F-distribution table for 5% significance level (link).

In order to find b so that P(W > b) = 0.05 one has to find the critical value for an $F_{3,10}$ distribution with 0.05 or 5% significance level. The values in the table will give you the threshold values below which 95% of the realisations of the random variable W fall (there are also F-distribution tables with 0.10 and 0.01 significance level).

For a F-distribution with 3 and 10 df. the critical value is 3.71 so that P(W > 3.71) = 0.05.